Boston College
Lynch School of Education and Human Development

Department of
Measurement, Evaluation, Statistics, and Assessment

MEASURING THE MATHEMATICS ANXIETY OF HIGH SCHOOL STUDENTS: AN APPLICATION OF RASCH MEASUREMENT THEORY

Dissertation by

KELSEY RUTH ERICKSEN KLEIN

Submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy

MAY 2021

ABSTRACT

MEASURING THE MATHEMATICS ANXIETY OF HIGH SCHOOL STUDENTS: AN APPLICATION OF RASCH MEASUREMENT THEORY

Kelsey R. E. Klein, Author

Larry H. Ludlow, Chair

The focus on science, technology, engineering, and mathematics fields has noticeably increased in recent decades with the rapid growth in technology. Mathematical knowledge and competence is in many ways a gateway to scientific and technical development, and therefore careers (Prieto & Delgado, 2007). Unfortunately, national and international assessments of mathematics achievement (e.g., Kena et al., 2016; NCES, 2015; OECD, 2013) show that schools in the United States are not helping students achieve acceptable levels of mathematical and numerical proficiency. Therefore, we need to understand how various factors, including mathematics anxiety, affect student performance and persistence in STEM. To do this, a valid and reliable measure of mathematics anxiety is needed. Existing instruments to measure mathematics anxiety have been insufficient in several ways, including in their conceptualization of the construct and use of classical test theory over Rasch measurement theory methods.

In this study, an instrument – the Comparative Mathematics Anxiety Scale (CMAS) – was developed to measure the three-dimensional conceptualization of mathematics anxiety that Dr. Caroline Vuilleumier and I originated. A unique, comparative item format adapted from Ludlow et al. (2014, 2019) and Rasch measurement theory (Rasch, 1960/1980) were utilized to mitigate some of the limitations of existing instruments. The overarching research question and three sub-research questions explored whether the CMAS could measure mathematics anxiety in a valid, reliable, and meaningful way.

This study employed a seven-step iterative scale development process and was accomplished across three rounds. Ultimately, twenty-three third-person items were developed to capture the emotional-attitudinal, mental-cognitive, and physical-somatic dimensions of mathematics anxiety. Using the Rasch rating scale model, the outcome was the 23-item CMAS that reliably and validly measures increasing levels of three dimensions of mathematics anxiety. The distribution of the items mostly confirmed their hypothesized order and the Rasch measurement theory principles. The scale also provides meaningful interpretations of what a raw score means regarding a student's experience of emotional-attitudinal, mental-cognitive, and physical-somatic mathematics anxiety.

Overall, the findings suggest that the novel approach of combining Rasch measurement theory with third-person items and comparative response options can be successful in developing a scale that measures an important construct. Furthermore, the scale can provide the evidence needed in the provision of interventions and in research to reduce students' overall experience of mathematics anxiety.

ACKNOWLEDGEMENTS

This PhD program was my life for 9 years, and I owe my ultimate success to so many. First, thank you to my committee. To Dr. Larry Ludlow, who has been with me from the beginning, thank you for your mentorship and for never giving up on me. It has been a long road, but we made it. To Dr. Laura O'Dwyer, who brought me back to Boston with an opportunity I couldn't refuse, thank you for pushing me to persevere and always reminding me of the end goal. To Dr. Lillie Albert, who didn't bat an eye at working with a group of methodologists, thank you for keeping students and mathematics education at the forefront.

Next, thank you to my colleagues at Education Development Center. To June Mark and Deborah Spencer, thank you for being the force of positivity that was so often needed. To Mary Beth Piecham, thank you for being a complete joy to work with and making the Institutional Review Board process painless (for me). To Julie Zeringue, Mary Fries, Kim Foster, Katie Chiappinelli, and Courtney Arthur, thank you for making work a place I want to be. Your inquiries into my dissertation progress were always thoughtful, encouraging, and appreciated.

To my family, thank you for being my anchor. No matter where I was at in the process or whether or not I was currently making progress, you all supported me unconditionally. To my mom, thank you for always being on speed dial (especially during comps), moving me too many times to count, and so much more. I will never be able to thank you enough.

To my friends, and to my fellow graduate students who became friends, thank you for making this process bearable and always being there for a laugh or nerdy moment when it was needed. You all made it worth it. A special shout out to the ERME Wives for always being a place for comfort, joy, and shenanigans—I needed you more than you'll ever know. And to the ERME Husbands, who unbeknownst to them joined a cult for life, thank you for rolling with the

punches and supporting all of us through this process. As the only "wife" without a "husband," you all cannot even begin to comprehend how much I appreciate and love all of you.

Last but most certainly not least, thank you to Dr. Caroline Vuilleumier and Mr. Derek Welch. Derek, thank you for welcoming me with open arms to your little family and providing me with a home away from home. I am forever grateful to you for not only putting up with Caroline throughout this process, but for accepting my constant presence without question. And to Caroline, who became my best friend and closest confidant during a 30-minute car ride to the Boston suburbs, this dissertation literally could not have happened without you. From convincing me to apply to the PhD program (yes, I blame you for this entire thing) to laying the foundation of this work with me, I owe it all to you. You said it best… You are my sister, and I love you.

ACKNOWLEDGEMENT OF NSF SUPPORT

This material is based upon work supported by the National Science Foundation under Grant No. 1621011. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

TABLE OF CONTENTS

Tables

Figures

CHAPTER 1 – INTRODUCTION

Overview of the Problem

Mathematics Education and Mathematics Anxiety

Mathematics and mathematics performance has long been a focus in education research. However, with the increase of technology in recent decades, the focus on science, technology, engineering, and mathematics (STEM) fields has increased substantially. Mathematical knowledge and competence is in many ways a gateway to scientific and technical development, and careers (Prieto & Delgado, 2007). Unfortunately, national and international assessments of mathematics achievement (e.g., Kena et al., 2016; NCES, 2015; OECD, 2013) show that schools in the United States are not helping students achieve acceptable levels of mathematical and numerical proficiency. Therefore, it has become increasingly important to understand the factors that affect performance. Various lines of research suggest that mathematics anxiety is a significant and prevalent impediment to mathematics achievement (Ashcraft & Moore, 2009).

Estimates of the prevalence of mathematics anxiety vary greatly, from 2-6% of secondary school students in England (Chinn, 2009) to about half of students enrolled in the most basic mathematics courses at Ohio State University in the United States (Betz, 1978). In England, 30% in a sample of apprentices had visible mathematics anxiety and 48% overall were affected significantly by mathematics anxiety (Johnston-Wilder, Brindley, & Dent, 2014). Notably, these estimates are likely to depend on the instrument used, the population being sampled, and how the cut-off for "high math anxiety" is defined (Dowker, Sarkar, & Looi, 2016). Regardless of the exact prevalence, mathematics anxiety appears to be a significant problem. And, while the exact prevalence of mathematics anxiety is uncertain, its relationship with mathematics performance and participation has been widely studied and corroborated.

As far as mathematics participation goes, people who experience mathematics anxiety have been found to avoid situations, classes, and careers that require using even basic mathematics (Brown, Brown, & Bibby, 2008; Chipman, Krantz, & Silver, 1992; Hopko, 2003; Richardson & Suinn, 1972). "Among nonstudents, mathematics anxiety may be a contributor to tensions during routine or everyday activities, such as handling money, balancing bank accounts, evaluating sale prices, or dividing work loads" (Richardson & Suinn, 1972, p. 551-552). These findings are of national import, for when the study of mathematics is avoided by otherwise capable students, opportunities to participate in advanced mathematics courses are curtailed and career options become more limited, which has the potential to erode a country's capital in science, technology, and other fields (Hembree, 1990).

Historically, "researchers have reported a consistent but small negative relationship between math anxiety and performance, with correlations generally ranging from about -.11 to -.36" (Ho et al., 2000, p. 362). More recent research suggests correlations on the higher end of the negative range with Prieto and Delgado (2007) obtaining a correlation of -0.29 between their measures of mathematics anxiety and mathematics achievement and Ashcraft & Kirk's (2001) correlations ranging from -0.27 to -0.34 in precollege and college samples. There are many hypotheses regarding the reason, or reasons, for the negative relationship between mathematics anxiety and mathematics performance, with reasons competing based on which direction the effect is theorized to occur; that is, whether high mathematics anxiety causes poor mathematics performance or whether poor mathematics performance causes high mathematics anxiety.

For example, one reason given is that individuals with higher levels of mathematics anxiety are more likely to avoid mathematical activities, therefore obtaining less experience, which in turn affects their performance on mathematics related tasks (Ashcraft, 2002). On the other hand, experiences of low mathematical performance may lead to higher levels of anxiety (Núñez-Peña & Suárez-Pellicioni, 2015). It is also possible to imagine how this process is cyclical: poor performance leads to an increase in levels of mathematics anxiety, which causes avoidance of mathematics activities and further low performance in mathematics, or moderate to high mathematics anxiety leading to poor mathematics performance, increasing one's mathematics anxiety and causing avoidance and further low performance in mathematics. Regardless of the specifics of the relationship, any relationship between mathematics anxiety and performance indicates that there is a confound in what we can say about a student's mathematics performance. That is, "any time a high-math anxious group performs more poorly than a low-anxious group, the interpretive question is whether their poor performance was due to high anxiety or to their lower level of mastery of the math" (Ashcraft & Moore, 2009, p. 201).

In addition, mathematics anxiety has been shown to be more prominent among subgroups of students. This may further exacerbate existing inequities in mathematics performance and participation, as well as opportunities to receive training in more advanced topics. For example, most research on gender differences suggests that female students experience greater mathematics anxiety than male students (Hembree, 1990). In terms of how gender and mathematics anxiety relate to mathematics performance, the relationship is unclear: some studies suggest no gender difference in the relationship between mathematics anxiety and performance or achievement (Meece, Wigfield, & Eccles, 1990; Ma, 1999; Wu, Barth, Amin, Malcarne, & Menon, 2012); Hembree's (1990) meta-analysis suggests that the negative relationship between mathematics anxiety and achievement is stronger for males than for females; and Devine, Fawcett, Szucs, and Dowker (2012) found a relationship between mathematics anxiety and performance in girls but not in boys after taking into account general test anxiety.

While gender is the most frequently studied demographic characteristic in the mathematics anxiety literature, socio-economic status, race/ethnicity, culture, and nationality may be other important factors. Unfortunately, little research has focused specifically on the relationship between mathematics anxiety and socioeconomic status; what research exists does not suggest a strong relationship between socioeconomic status and mathematics anxiety (Jadjewski, 2011). On the other hand, race/ethnicity, culture, and nationality has been explored a bit more with some interesting results. Research in the United States (Catsambis, 1994; Lubienski, 2002) and the United Kingdom (National Audit Office, 2008) suggests white students may have higher mathematics anxiety than ethnic minority students in these countries. Dowker et al. (2016) found students in high-achieving Asian countries showed higher mathematics anxiety than students in high-achieving Western European countries. Both of these findings indicate the culture of these racial/ethnic groups and countries may influence students' mathematics anxiety. Even with these indications, more research is needed to corroborate these results and investigate the impact of nationality, race/ethnicity, and culture on the relationship between mathematics anxiety and performance. Understanding mathematics anxiety and its relationship to student demographics separately, and in combination with their mathematics performance and participation, allows for the potential to address mathematics anxiety. This in turn may help tackle the achievement gap and contribute to broadening participation in mathematics and other STEM disciplines.

Defining Mathematics Anxiety

An important step to understanding mathematics anxiety and its relationship to other constructs is in appropriately defining and then measuring the construct. The most commonly cited definition of mathematics anxiety is that put forth by Richardson and Suinn (1972), which states: “mathematics anxiety involves feelings of tension and anxiety that interfere with the manipulation of numbers and the solving of mathematical problems in a wide variety of ordinary life and academic situations” (p. 551). However, this definition of mathematics anxiety is very simplistic and does not take into account the potential multi-dimensional nature of mathematics anxiety (Baloglu & Zelhart, 2007; Bandalos, Yates, & Thorndike-Christ, 1995; Kazelskis, 1998; Meece et al., 1990; Wigfield & Meece, 1988). Therefore, other definitions of mathematics anxiety have been explored, although not to a great extent. It is important to note that there has been little consistency in how mathematics anxiety has been defined, as researchers often fail to put forth explicit definitions of the construct they aim to measure or borrow instruments from other disciplines without appropriately re-defining the instrument for use in mathematics. This has resulted in confusion on the part of readers and other researchers (Ma & Kishor, 1997). “These uncertainties and contradictions on how mathematics anxiety is conceptualised have resulted in no commonly accepted construct model and the need to refine measures” (Cavanagh & Sparrow, 2011, p. 166).

As a response to this finding, Cavanagh and Sparrow (2011), building off the work of others, proposed a three-dimensional definition of mathematics anxiety. The three dimensions they proposed are somatic (physical and body), cognitive (mental processes), and attitudinal (emotional) indicators of mathematics anxiety. Further, they defined these dimensions as levels, by which somatic anxiety is extremely high anxiety, cognitive anxiety is high (or moderate) anxiety, and attitudinal anxiety is low anxiety. However, after the development of an instrument created with this construct model and the analysis of data collected with this instrument, Cavanagh and Sparrow (2011) found “that high anxiety, as well as low anxiety, is indicated by a combination of attitudinal, cognitive, and somatic indicators” (p. 172). Therefore, they revised their conceptualization of mathematics anxiety to include somatic, cognitive, and attitudinal mathematics anxiety as separate dimensions that all span from low to high anxiety. Cavanagh and Sparrow's (2011) revised conceptualization served as the starting point for the theoretical framework that was used and ultimately revised in this dissertation.

Existing Measures of Mathematics Anxiety and their Limitations

In addition to the instrument developed by Cavanagh and Sparrow (2011), a number of other instruments have been created to measure mathematics anxiety (e.g., Fennema & Sherman, 1976; Prieto & Delgado, 2007; Richardson & Suinn, 1972; Wigfield & Meece, 1988). The most frequently used, cited, and adapted of these instruments is the Mathematics Anxiety Rating Scale (MARS), developed by Richardson and Suinn (1972) and described in detail in the literature review in Chapter Two, given its availability and accompanying psychometric data. Even though this scale has been widely used, several limitations of the scale exist. First, there is no clear conceptualization of mathematics anxiety underlying the item development. Second, the scale does not appear to take into account the multi-dimensional nature of mathematics anxiety. Finally, the scale was developed using classical test theory principles and ideas, which as will be discussed in the literature review has some limitations compared to other theories, such as the Rasch measurement theory.

Other instruments developed to measure mathematics anxiety have avoided some of the limitations of Richardson and Suinn's (1972) scale. For example, Wigfield and Meece (1988) developed the Mathematics Anxiety Questionnaire (MAQ) by building off of Liebert and Morris' (1967) two-dimensional conceptualization of test anxiety, which includes both affective and cognitive dimensions. This conceptualization of mathematics anxiety takes into account the construct's multi-dimensional nature but still does so from the classical test theory perspective.

Prieto and Delgado (2007), on the other hand, did not take into account the multi-dimensional nature of mathematics anxiety but used the Rasch rating scale model (Andrich, 1978; Wright & Masters, 1982) in their analysis. It is important to note, however, Prieto and Delgado (2007) did not explicitly develop their instrument using the ideas and principles of Rasch measurement theory, at least not as described. That is, they used the statistical methods associated with a Rasch model analysis but without also applying the theoretical basis for the method. To date, from my review of existing instruments, Cavanagh and Sparrow (2011) are the first to take into account the multi-dimensional nature of mathematics anxiety, even including a third dimension to measure somatic mathematics anxiety, while also applying some Rasch measurement theory principles and using Rasch analysis methods. Even as such, Cavanagh and Sparrow (2011) have noted the need for adjustments to their conceptualization of mathematics anxiety. Therefore, a psychometrically robust instrument that takes advantage of Rasch measurement theory is needed to advance the field; the beginning of this work is described in the following section.

Recent Work to Measurement Mathematics Anxiety

Based on the limitations in the conceptualization of mathematics anxiety and the development of instruments used to measure mathematics anxiety, another doctoral student and I wrote items for our Mathematics Testing Anxiety Scale (Vuilleumier & Klein, 2015). Prior to developing the items for this scale, we developed a theoretical framework based on Cavanagh and Sparrow's revised understanding of mathematics anxiety, where they theorized that somatic, cognitive, and attitudinal mathematics anxiety are not separate levels but rather different dimensions of mathematics anxiety that *all* can range from high to low anxiety (*Table 1*, on the following page).

Table 1

Original Mathematics Anxiety Theoretical Continuum (Fall 2013 – Spring 2015)

Level of Anxiety	Dimension of Mathematics Anxiety		
	Emotional/Attitudinal	*Mental/Cognitive*	*Physical/Somatic*
Overwrought	Feeling overwhelmed, fearful, or threatened by the task	"Blanking out" or absent-mindedness in which one cannot recall what was studied or learned	Racing heartbeat, difficulty breathing, or physically shaking
Distressed	Feeling concerned, worried, or distressed about/by the task	Confusion, frustration, and feeling like one should know what to do but cannot put one's thoughts in order	Stomachache, dry mouth, or sweaty palms
Antsy	Feeling uneasy or unsettled about the task	Mental fogginess or scatterbrained thoughts in which it takes considerable effort to gather one's thoughts in a logical way	Nervous ticks such as tapping one's foot, squirming in one's seat, or biting one's nails
Apprehensive	Feeling nervous or hesitant about the task	Inability to focus or being easily distracted	Muscle stiffness or tension headaches
Tranquil	No indicators of anxiety; unfazed	No indicators of anxiety; focused	No indicators of anxiety; appears calm

Using our new framework, twenty-five items spanning five levels were written to measure mathematics testing anxiety (*Table 2*, on the following page). Two rounds of data collection were conducted with undergraduate students enrolled in at least one mathematics class and analyses were run to determine whether or not the theorized conceptualization of mathematics testing anxiety was supported by the data.

Table 2

Original Comparative Mathematics Testing Anxiety Items

Level of Anxiety	Dimension of Mathematics Anxiety		
	Emotional/Attitudinal	*Mental/Cognitive*	*Physical/Somatic*
Overwrought	1_6. Jason is feeling overwhelmed.	2_1. Emily's mind keeps blanking out. 2_4. Josh can't recall what he learned for the exam.	3_2. Alex is on the verge of tears.
Distressed	1_1. Pete is distressed. 1_3. Liz is worried.	2_3. Joe feels like he should know how to solve the problem but he can't put his thoughts in order. 2_5. Maggie feels confused.	3_6. Mike has a stomachache. 3_8. Rachel's heart is racing.
Antsy	1_5. Ruth is feeling uneasy. 1_8. Brendan is on edge.	2_8. Ryan's thoughts are scattered. 2_9. Sarah has a mental fog.	3_1. Abby is biting her nails. 3_7. Phil is squirming in his seat.
Apprehensive	1_2. Michelle is feeling apprehensive. 1_4. Will is feeling timid.	2_6. Tammy is easily distracted while taking her exam.	3_3. Beth is feeling tense. 3_5. James feels physically uncomfortable.
Tranquil	1_7. Valerie is feeling unfazed.	2_2. Grace is focused on the exam. 2_7. Matt is thinking clearly.	3_4. Dan appears calm.

Unfortunately, our study did not support the theorized conceptualization of mathematics testing anxiety (the results of this research will be presented further in the literature review), and therefore presumably did not support the theorized conceptualization of general mathematics anxiety. Generally, the analyses showed that while the somatic, cognitive, and attitudinal dimensions of mathematics anxiety spanned multiple levels (as the theory suggested), the dimensions were also leveled such that items measuring somatic mathematics anxiety spanned the highest levels, items measuring cognitive mathematics anxiety spanned the middle levels, and items measuring attitudinal mathematics anxiety spanned the lowest levels. Therefore, the individual dimensions appear to span multiple levels of mathematics anxiety but not all levels of mathematics anxiety. This result could be an artifact of poor item writing such that items simply were not appropriately written to span across all levels of the construct. On the other hand, given Cavanagh and Sparrow's (2011) similar results, an adjustment to the theory, such that the three dimensions are leveled, as they originally suggested, but with some overlap may be warranted. Therefore, a more modest change to the theory, rather than jumping to all three dimensions spanning all levels of mathematics anxiety, may have been more appropriate.

It is important to note that this work came from a single study, although across two rounds of data collection, and data was collected from undergraduate students at a single institution, instead of from high school students. Therefore, additional data collection is likely warranted before changes are made to the conceptualization of mathematics anxiety. In addition, the items were written to measure mathematics *testing* anxiety instead of general mathematics anxiety, providing an additional reason for revision and testing prior to making adjustments to the theory of mathematics anxiety. While the scale was intended to measure mathematics anxiety in a testing scenario, the items should be easily adapted to measure general mathematics anxiety, and I theorize these edits would not substantially change the behavior of the items. Revision to the theoretical construct of mathematics anxiety as well as revisions to the instrument itself is the work that this dissertation aims to accomplish.

Purpose of the Proposed Study

The purpose of the proposed study is to develop and refine a measure of mathematics anxiety for use with high school students that is psychometrically robust and overcomes the limitations of existing instruments alluded to earlier. A psychometrically robust instrument is one that is consistent across samples and maintains its psychometric properties across different contexts. My prior theoretical framework and instrument development work with Dr. Caroline Vuilleumier will provide the foundation for the development of an instrument with high construct validity. Based on the purpose of the study and existing theories of mathematics anxiety, the broad research question of interest is: To what extent can the construct of mathematics anxiety be measured reliably and meaningfully by developing a Rasch-based scale? This question will be answered by answering the following sub-questions:

1. How can the tripartite conceptual framework of mathematics anxiety Dr. Caroline Vuilleumier and I developed be employed and adapted in the development of a Rasch-based scale?
2. Is the scale invariant across grade, gender, and other student demographics?
3. Does the scale have an acceptable degree of construct and criterion validity?

Significance of the Study

This dissertation aims to fill a void in the mathematics anxiety measurement field by clarifying our conceptual understanding of mathematics anxiety while developing an instrument to measure the construct. Existing instruments to measure mathematics anxiety are outdated, often developed without a clear framework of mathematics anxiety, and most frequently rely on the principles and methods of classical test theory rather than Rasch measurement theory (which poses its own set of limitations). A psychometrically robust instrument to measure mathematics anxiety that is developed with clear, substantiated conceptual underpinnings is needed to advance the field and its ability to appropriately measure students' levels of mathematics anxiety. Such a measure would allow researchers to more accurately measure changes in mathematics anxiety due to mathematics interventions and more accurately depict the relationship between mathematics anxiety and other important constructs such as mathematics achievement, performance, and participation. Through the course of this research, I will also be contributing to an NSF-funded study, *Supporting Success in Algebra* [NSF 1621011], which aims to look at the impact of an algebra-support curriculum intervention on student mathematics achievement and attitudes towards mathematics, including their anxiety in mathematics.

Summary

This chapter introduced the problem of mathematics anxiety and its impact on student achievement and learning in mathematics, as well as students' likelihood to avoid pursuing careers that require even basic skills in mathematics. Of even greater concern is that mathematics anxiety has been found to be more prevalent among some subgroups of student (for example, female students). In addition, this chapter highlighted the persistent problem in defining mathematics anxiety and therefore the constant lack of clarity and need to refine measures, as well as the reliance of most measures on classical test theory principles and methods. Therefore, the need exists for an instrument to measure mathematics anxiety that is based on a clearly-defined conceptualization of mathematics anxiety and utilizes the principles and methods of Rasch measurement theory. The purpose of this study is to develop such an instrument. Finally, I discuss the contributions of this study to the fields of psychology and mathematics education. In the following literature review chapter, I present current theories of mathematics anxiety, research showing mathematics anxiety's relationship to other constructs, and existing instruments that measure mathematics anxiety. I also discuss Rasch measurement theory and its advantages over other methods, as well as the instrument development work that was completed prior to this dissertation study.

CHAPTER 2 – LITERATURE REVIEW

This chapter will review the extant literature to justify and inform the proposed dissertation research. First, prevailing theories of mathematics anxiety will be reviewed. Then, I will describe the relationships between mathematics anxiety and related constructs found in the literature. This will be followed by a review of existing instruments that were designed to measure mathematics anxiety among high school students. Next, the Rasch measurement theory approach guiding this study's instrument development and data analysis will be discussed including its advantages over classical test theory. Finally, I will describe the initial instrument development efforts of this research that occurred between September 2013 and April 2015.

Theory of Mathematics Anxiety

Defining Mathematics Anxiety

As stated previously, the most commonly cited definition of mathematics anxiety is "mathematics anxiety involves feelings of tension and anxiety that interfere with the manipulation of numbers and the solving of mathematical problems in a wide variety of ordinary life and academic situations" (Richardson & Suinn, 1972, p. 551). Fennema and Sherman's (1976) definition of mathematics anxiety, "feelings of anxiety, dread, nervousness, and associated bodily symptoms related to doing mathematics" (p. 326), suggests a potentially multi-dimensional construct as it includes physical reactions (associated bodily symptoms). Another interesting aspect of Fennema and Sherman's (1976) definition is that they propose anxiety in mathematics as ranging "from feeling at ease to feeling distinct anxiety" (p. 326), which suggests that mathematics anxiety exists along a continuum for which they have defined the extremes.

Other definitions of mathematics anxiety take the dimensionality of the construct even further by suggesting indicators that can be classified into three unique categories: psychophysiological (Faust, 1992; Lang, 1968), cognitive (Lang, 1968; Richardson & Woolfolk, 1980), and behavioral (Lang, 1968). Haase et al. (2019) describe mathematics anxiety as having four levels (or dimensions): cognitive, affective, behavioral, and physiological. Building off of some of this and other previous work, Cavanagh and Sparrow (2011) proposed a three-dimensional definition of mathematics anxiety containing somatic, cognitive, and attitudinal indicators. Somatic indicators of mathematics anxiety include sweaty palms, nausea, difficulty breathing, and heart palpitations (Cavanagh & Sparrow, 2011; Malinski, Ross, Pannells, & McJunkin, 2006; Perry, 2004). Cognitively, mathematics anxiety interferes with calculating and solving mathematical problems in a variety of contexts (Cavanagh & Sparrow, 2011; Richardson & Suinn, 1972; Suinn, Taylor & Edwards, 1998). Finally, psychological, emotional, attitudinal, or affective indicators include, for example, feelings of tension, fear, apprehension, dysphoria, and feeling threatened (Ashcraft & Kirk, 2001; Cavanagh & Sparrow, 2011; D'Ailly & Bergering, 1992; Jain & Dowson, 2009; Perry, 2004; Richardson & Suinn, 1972; Zohar, 1998). If one were to consider the fourth dimension, behavior, indicators would include avoiding engagement in mathematics tasks, issues of concentration, and inability to follow instructions in math.

A key feature of Cavanagh and Sparrow's (2011) conceptualization of mathematics anxiety is their recognition that latent traits have a developmental component that acknowledges the potential for human change and growth. Even with this assumption, the literature has been relatively sparse on the developmental nature of mathematics anxiety; that is, how mathematics anxiety develops from low, or even non-existent, to overwhelming. One study that does not fit this mold is Prieto and Delgado (2007) where they developed a model of mathematics anxiety such that "nausea" indicates a higher level of anxiety and "one's mind going blank" indicates a lower level of anxiety. Cavanagh and Sparrow (2011) extend the hierarchical component of mathematics anxiety, which they call the developmental assumption, by hypothesizing that the three dimensions of mathematics anxiety (somatic, cognitive, and attitudinal) are actually levels, by which somatic anxiety is high anxiety, cognitive anxiety is moderate anxiety, and attitudinal anxiety is low anxiety. Here, somatic indicators include nausea and heart palpitations; cognitive indicators include confusion and one's mind going blank; and attitudinal anxiety indicates low mathematical confidence. They also suggest that these indicators are cumulative such that a person showing high anxiety will exhibit the low anxiety indicators as well. However, these lower indicators may not be so obvious as they are overshadowed by the indicators of higher levels of anxiety. Their model also acknowledges the various contexts and scenarios that may result in mathematics anxiety.

Differentiating between Mathematics Anxiety, Test Anxiety, and General Anxiety

In defining mathematics anxiety, it is important to consider the construct's relationship to both test anxiety and general anxiety. Previous studies suggest that mathematics anxiety is closely related to other measures of anxiety, such as text anxiety and general anxiety, and may be more related to these measures than to measures of academic achievement (Ashcraft, Kirk, & Hopko, 1998; Hembree, 1990). Ashcraft and Moore (2009) found that mathematics anxiety correlates most strongly with test anxiety but is also moderately correlated with general, trait, and state anxiety. Dowker et al. (2016) even suggest that general anxiety may explain some of the correlation, or relationship, between mathematics anxiety and test anxiety; that is, general anxiety may mediate the relationship between mathematics anxiety and test anxiety. It is important to note that previous research suggests that mathematics anxiety is a unique construct given that it correlates only moderately with general anxiety and test anxiety and occurs only in very specific situations (Prieto & Delgado, 2007).

Much of the research on the relationship between mathematics anxiety and other measures of anxiety focus specifically on the relationship between test anxiety and mathematics anxiety. In this research, it has often been suggested that mathematics anxiety is a subject-specific display of test anxiety (e.g., Bandalos et al., 1995; Brush, 1981; Dew, Galassi, & Galassi, 1983; Hembree, 1990; Ho et al., 2000). Under this proposal, it could be presumed that the theoretical models of test anxiety could also be used for mathematics anxiety. The theoretical models of test anxiety can be separated into two broad categories: those that address the construct's dimensionality and those that address the direction of effects. For direction of effects, the inference model postulates that test anxiety interferes with students' ability to recall information learned previously, while the deficit model proposes that poor performance leads to higher levels of test anxiety (Hembree, 1990). While this debate has not been completely put to rest, Hembree's (1988) meta-analysis concluded that test anxiety causes poor performance, supporting the inference model. As far as the dimensionality of test anxiety, it was originally theorized to be a one-dimensional construct; however, in 1967, Liebert and Morris suggested test anxiety was actually a two-dimensional model comprised of an affective and a cognitive component. A discussion of the dimensionality of mathematics anxiety follows.

The Dimensionality of Mathematics Anxiety

One aspect of the conversation that deserves more focused discussion is on the dimensionality of mathematics anxiety. Some have found the construct to be multi-dimensional (Baloglu & Zelhart, 2007; Bandalos et al., 1995; Haase et al., 2019; Hopko, 2003; Meece et al., 1990; Wigfield & Meece, 1988) while others suggest it is a one-dimensional construct (Beasley, Long, & Natali, 2001). Wigfield and Meece (1988), specifically, found two separate dimensions of mathematics anxiety, cognitive and affective, in their study of sixth graders and secondary school students. These are similar to the dimensions previously identified in the test anxiety literature by Liebert and Morris (1967). The cognitive dimension, "worry," refers to concern about one's performance and the consequences of failure, while the affective dimension, "emotionality," refers to nervousness and tension in testing situations and respective autonomic reactions (Liebert & Morris, 1967. As mentioned previously, Haase et al. (2019) describe mathematics anxiety as having four dimensions: cognitive, affective, behavioral, and physiological. While few in number, these studies suggest mathematics anxiety is a multi-dimensional construct, and the different dimensions of the construct can vary in significant ways.

Mathematics Anxiety and Its Relationship to Other Constructs

This section aims to provide a broad summary of what is known about mathematics anxiety and its relationship to other constructs. Specifically, this section will focus on the relationship between mathematics anxiety and the following student attributes: demographics, genetics, mathematics achievement, attitudes towards mathematics, and future plans in mathematics, as well as the influence of others on individuals' mathematics anxiety.

Demographics

Student demographics are often included as covariates or variables of interest in research on student achievement, attitudes, beliefs, and more; research on mathematics anxiety is no different. This section details some of the literature on the relationship between student demographics, namely student gender, age, socioeconomic status, race/ethnicity, culture, and nationality, and student mathematics anxiety.

While research on the relationship between mathematics anxiety and gender is not completely consistent (Baloglu & Koçak, 2006), a majority of research has found that female students exhibit greater mathematics anxiety than their male counterparts, although the effect size may be small (Devine et al., 2012; Else-Quest, Hyde, & Linn, 2010; Hembree, 1990; Hyde, 2005; Hyde, Fennema, Ryan, Frost, & Hopp, 1990; Prieto & Delgado, 2007; Wigfield & Meece, 1988). This gender difference has been found even in countries where there is little to no difference in students' actual mathematics performance (Spelke, 2005). Potential reasons for female students' increased mathematics anxiety include lower self-ratings in mathematics, general differences in levels of anxiety and confidence for male and female students, exposure to gender stereotypes in mathematics, and transmission of mathematics anxiety by female teachers to their female students (Beilock, Rydell, & McConnell, 2007; Beilock, Gunderson, Ramirez, & Levine, 2010; Dowker et al., 2016; Johns, Schmader, & Martens, 2005; Schmader, 2002).

Research on gender and mathematics anxiety frequently has looked at the impact of gender on the relationship between mathematics anxiety and mathematics performance. Some studies suggest no gender difference (Meece et al., 1990; Ma, 1999; Wu et al., 2012); one meta-analysis suggests the negative relationship between the constructs is stronger for males than for females (Hembree, 1990); and Devine et al. (2012) found the relationship existed in girls but not in boys after accounting for general test anxiety. Dowker et al. (2016) note that studies finding no difference or a stronger negative relationship for boys than for girls did not take into account general test anxiety. "Gender effects on the relationship between mathematics anxiety and performance may also depend on whether one is examining the cognitive or affective component of mathematics anxiety, and on what aspects of mathematics are involved" (Dowker et al., 2016, p. 8). In their cross-national study on the affective and cognitive components of mathematics anxiety in 6^{th} grade students, Ho et al. (2000) found no gender differences in affective mathematics anxiety for China and the Unites States and in cognitive mathematics anxiety for China. However, they found female students had higher affective mathematics anxiety than male students in Taiwan and female students had higher cognitive mathematics anxiety than male students in both Taiwan and the United States. Related to the second point, Miller and Bichsel (2004) found mathematics anxiety was more strongly related to applied mathematics scores for female students but to basic mathematics scores for male students.

In addition to gender, age has also been found to have a relationship with students' mathematics anxiety, or perception of mathematics anxiety (Furner & Berman, 2003; Hembree, 1990; Jackson & Leffingwell, 1999). Generally speaking, mathematics anxiety appears to increase with age during childhood (Dowker et al., 2016; Haase, et al., 2019; Hembree, 1990). This finding is consistent with other findings suggesting students' attitudes towards mathematics deteriorate as they age (Blatchford, 1996; Dowker, 2005; Ma & Kishor, 1997; Mata, Monteiro, & Peixoto, 2012; Wigfield & Meece, 1988). Viable reasons for this arguable increase in levels of mathematics anxiety with age include an increase in general anxiety with age, exposure to social stereotypes toward mathematics, exposure to failure or the threat of failure in mathematics, and changes in the mathematics content learned (Dowker et al., 2016). Note, little research has focused specifically on the relationship between mathematics anxiety and socioeconomic status; what research exists does not suggest a strong relationship between socioeconomic status and mathematics anxiety (Jadjewski, 2011).

Race/ethnicity, culture, and nationality are other demographic characteristics of students that may impact their mathematics anxiety. Interestingly, within the United States (Catsambis, 1994; Lubienski, 2002) and the United Kingdom (National Audit Office, 2008), research suggests that ethnic minority students have more positive attitudes towards mathematics, including lower mathematics anxiety, than their white counterparts. Dowker et al. (2016) found that children in high-achieving Asian countries tended to demonstrate higher levels of mathematics anxiety than those in high-achieving Western European countries, though the reasons for these differences were unclear. Stankov (2010) provides evidence that it may be the high-stakes academic culture experienced by Asian students that both influences students to perform well but also increases their anxiety and self-doubt.

A Genetic Component to Mathematics Anxiety?

In one of the few genetic studies on mathematics anxiety, Wang et al. (2014) used behavioral genetic modeling to look at the influence of genetics on mathematics anxiety for a sample of 514 pairs of twelve-year-old twins. They found a significant positive correlation between mathematics anxiety and general anxiety, as well as significant negative correlations between mathematics anxiety and both mathematical problem solving and reading comprehension. Interestingly, general anxiety did not have a significant correlation with either mathematical problem solving or reading comprehension. Wang et al. (2014) also found "that genetic factors accounted for about 40% of the variance in mathematics anxiety, with most of the rest being explained by non-shared environmental factors… Thus, mathematics anxiety may result from a combination of negative experiences with mathematics, and predisposing genetic risk factors associated with both mathematical cognition and general anxiety" (Dowker et al., 2016, p. 7). Malanchini et al. (2017) obtained similar results, indicating a role for both shared and non-shared genetic influences as well as non-shared environmental factors.

Mathematics Anxiety and Mathematics Ability, Achievement, and Performance

While it has been found that mathematics anxiety is not correlated to intelligence (Ashcraft & Moore, 2009), most research suggests a moderate to strong negative relationship between mathematics anxiety and mathematics achievement or performance (e.g., Foley et al., 2017; Hembree, 1990; Liebert & Morris, 1967; OECD, 2013; Richardson & Suinn, 1972;

Sarason, 1986; Wigfield & Meece, 1988), with the range of correlations typically between -0.11 and -0.36 (Ho et al., 2000). There is also evidence to suggest that the relationship between mathematics anxiety and achievement is not mediated either by verbal aptitude and achievement or by mathematics competence (Dowker et al., 2016; Hembree, 1990). Sherman and Fennema (1978) argue that mathematics anxiety may be the most important attitudinal variable in predicting mathematics achievement; others agree that mathematics anxiety plays a distinct and large role in mathematics performance (Baloglu & Koçak, 2006; Ho et al., 2000; Ma & Kishor, 1997; McLeod, 1992; Miller & Bichsel, 2004).

There are many hypotheses regarding the reason for the negative relationship between mathematics anxiety and mathematics achievement or performance, with some of the hypothesized reasons discussed earlier. There is an additional, more direct argument, for the influence of mathematics anxiety on mathematics performance, which is that mathematics anxiety overloads an individual's working memory (Ashcraft et al., 1998). High levels of mathematics anxiety have been shown in some experimental studies to impair working memory and math problem solving (Hopko, Ashcraft, Gute, Ruggiero, & Lewis, 1998; Kellogg, Hopko & Ashcraft, 1999; Ashcraft & Kirk, 2001; Hopko, McNeil, Lejuez, Ashcraft, Eifert, & Reil, 2003). A general explanation for this is given by processing efficiency theory (Eysenck & Calvo, 1992), which posits that working memory dysfunction is caused when anxious individuals use space in their working memory to focus on their negative feelings and beliefs rather than on the task thus reducing their ability to perform (Ferrando, Varea, & Lorenzo, 1999). According to Ashcraft and Kirk (2001), individuals with higher levels of mathematics anxiety have shorter working memory spans than individuals with lower mathematics anxiety, which can lead to a drop in mathematics performance. Ashcraft and Moore (2009) have found evidence for this drop in performance that can be attributed to mathematics anxiety, which they termed an *affective drop*.

One final area in the literature on the relationship between mathematics anxiety and mathematics performance worth further review is that which accounts for the different components of mathematics anxiety, namely affective and cognitive mathematics anxiety. Research suggests that the affective and cognitive components of mathematics anxiety are differentially related to mathematics achievement and performance. Going back to test anxiety (and the cognitive and affective components of test anxiety), research consistently shows a significant negative relationship between test performance and cognitive test anxiety, but the relationship between affective test anxiety and test performance has been less consistent (Deffenbacher, 1980; Liebert & Morris, 1967; Morris, Davis, & Hutchings, 1981; Tyron, 1980; Wine, 1971). In contrast, for measures of mathematics anxiety, it appears it is the affective dimension that exhibits significant negative correlations with mathematics performance more strongly than the cognitive dimension (Ho et al., 2000; Wigfield & Meece, 1988). Specifically, for samples in China, Taiwan, and the Unites States, Ho et al. (2000) found a consistent and significant negative relationship between mathematics achievement and the affective dimension of mathematics anxiety. However, the relationship between the cognitive dimension of mathematics anxiety and mathematics achievement varied across the three national samples, with the samples from the United States and China exhibiting an insignificant relationship, and the Taiwanese sample exhibiting a significant, positive correlation.

Mathematics Anxiety and Attitudes towards Mathematics

Mathematics anxiety has frequently been found to have negative relationships with students' attitudes towards mathematics. For example, D'Ailly and Bergering (1992) found a significant, although small, correlation between mathematics anxiety and mathematics avoidance. In precollege samples, Ashcraft and Moore (2009) found negative correlations, ranging from -0.37 to -0.82, between mathematics anxiety and the following constructs: usefulness of mathematics, ratings of mathematics teachers, motivation in mathematics, enjoyment of mathematics, and self-confidence in mathematics. These findings are similar to those shown by Hembree (1990) where mathematics anxiety averaged correlations of -0.73 and -0.82 with enjoyment of mathematics and confidence in mathematics, respectively. Perhaps most prominently, mathematics anxiety shows negative correlations to measures of the related constructs of mathematics self-concept, mathematics self-rating, and mathematics self-efficacy (Cooper & Robinson, 1991; Goetz, Cronjaeger, Frenzei, Ludtke, & Hall, 2010; Hoffman, 2010; Jain & Dowson, 2009; Lee, 2009; Pajares & Miller, 1994).

Mathematics Anxiety's Relationship to Coursework and Future Plans

In addition to being related to some of the previously described characteristics, mathematics anxiety has also been found to be correlated to both mathematics coursework (current enrollment in as well as future plans to take) and college major. Ashcraft and Moore (2009) found that mathematics anxiety "correlates -.31 with the extent of high school math courses taken (electives) and -.32 with intent to enroll in more math courses" (p. 200). As far as college major, Hembree (1990) found, perhaps predictably, that college students with low mathematics anxiety were more likely to major in mathematics, the physical sciences, and engineering, whereas college students with higher levels of mathematics anxiety were more likely to major in the humanities and other non-math-related disciplines. Perhaps most worrisome, Hembree (1990) found elementary education majors had the highest level of mathematics anxiety.

Influence of Others on Mathematics Anxiety

Teachers, parents, and peers have the ability to impact an individual's level of mathematics anxiety, both positively and negatively. Overall, teachers, particularly elementary teachers, have high levels of mathematics anxiety (Beilock et al., 2010; Hembree, 1990), and this anxiety can be passed on to students. Of particular concern, there is evidence that female teachers' mathematics anxiety can be transmitted to female students (Beilock et al., 2010). Other ways in which teachers can negatively influence students' mathematics anxiety include social exposure to math failure in the classroom (Ashcraft, Krause, & Hopko, 2007) and negative learning experiences where there is high demand for correctness with little motivational and cognitive support (Bekdemir, 2010; Meece, Wigfield, & Eccles, 1990; Turner et al., 2002).

Parents and families exert influence on students' levels of mathematics anxiety through attachment, expectancy socialization, over-control or punishment, reinforcement, and role models (Batchelor et al., 2017). Additionally, students are particularly affected by their peers. Students' development of a math identity as well as their attitudes towards mathematics are influenced by their peers' recognition and attitudes. Finally, while classrooms segregated by performance level are beneficial for high-performing students, there is a negative effect for lower-performing students (Frenzel, Goetz, Pekrun, & Watt, 2010). Thus, in a variety of ways, peers, families, and teachers all have the potential to increase students' mathematics anxiety.

Existing Instruments to Measure Mathematics Anxiety

Thus far, we have reviewed the literature on mathematics anxiety and its relationship to other constructs without reference to the measures used to study them. However, the methods and measures used to study mathematics anxiety and its related constructs are important, as they are the means by which we are able to study these concepts. How concepts are defined and ultimately measured can have a major impact on research results. For example, when assessing the discrepancies between Spielberger's (1977) Test Anxiety Inventory and Wigfield and Meece's (1988) Math Anxiety Questionnaire, both of which include an affective and a cognitive scale, Williams (1994) found the affective scales and cognitive scales of both scales shared approximately 24% and 13% of their variance, respectively. While the two scales measure different constructs, test anxiety and mathematics anxiety, one might assume that the affective and cognitive components would be more highly related. This would indicate a lack of convergent validity between the two measures, meaning the two instruments likely defined affective and cognitive anxiety differently. This example shows the importance of construct definition and how varying definitions can lead to scales measuring vastly different concepts.

A majority of mathematics anxiety measures involve questionnaires and rating scales, which have been predominately tested with adolescents and adults (Dowker et al., 2016). To my knowledge and others' (Ashcraft & Moore, 2009; Dowker et al., 2016), Dreger and Aiken (1957) developed the first known questionnaire or objective measure of 'mathematics anxiety,' which they called more generally 'numerical anxiety.' Their instrument, the Numerical Anxiety Scale, took the Taylor Manifest Anxiety Scale (Taylor, 1953) and added three math-related items. Since Dreger and Aiken's (1957) development of the Numerical Anxiety Scale, two mathematics anxiety instruments have been developed and widely used: the Mathematics Anxiety Research Scale, or MARS (Richardson & Suinn, 1972), and the Anxiety subscale of the Fennema-Sherman Mathematics Attitude Scales (Fennema & Sherman, 1976). "The MARS is a 98-item scale composed of brief descriptions of behavioral situations… that may arouse different levels of anxiety" (Richardson & Suinn, 1972, p. 552). Participants are asked to rate themselves on a scale from 1, *not at all anxious*, to 5, *very much anxious*, where high scores represent high mathematics anxiety. The Anxiety subscale of the Fennema-Sherman Mathematics Attitude Scale is a 12-item scale that "is intended to measure feelings of anxiety, dread, nervousness, and associated bodily symptoms related to doing mathematics. The dimension ranges from feeling at ease to feeling distinct anxiety. The scale is not intended to measure confidence in, or enjoyment of, mathematics" (Fennema & Sherman, 1976, p. 326). Participants are asked to rate themselves on a scale from 1, *strongly disagree*, to 5, *strongly agree*. While Fennema and Sherman's (1976) Anxiety subscale is shorter, Richardson and Suinn's (1972) MARS has been more widely used and adapted given its relative ease of access and large range of psychometric data; therefore, further attention will be given to this instrument below.

Richardson and Suinn (1972), Mathematics Anxiety Rating Scale (MARS)

As stated before, the original MARS developed by Richardson and Suinn (1972) is a 98-item scale with five response options from *not at all anxious* to *very much anxious*. The scale was found to have adequate test-retest reliability of 0.85 and high internal consistency of 0.97 (Richardson & Suinn, 1972). They also found validity evidence in two forms: 1) after participating in behavior therapy aimed at reducing mathematics anxiety, scores on the MARS appropriately decreased, and 2) a negative correlation was found between MARS scores and performance on a mathematics exam.

However, many have found the length of the original MARS burdensome and unnecessary for maintaining adequate psychometric properties; therefore, several offspring of the MARS have been created that are reduced in length while maintaining good psychometric properties. Examples that are appropriate for high school students include Ashcraft and Kirk's (2001) 25-item sMARS (shortened MARS), which is based on a version created by Alexander and Martray (1989) and correlates 0.97 with the original MARS; and Hopko, Mahadevan, Bare, and Hunt's (2003) 9-item AMAS (Abbreviated Math Anxiety Scale). Versions have also been created for use with younger students (e.g., MARS-A, adolescents, and MARS-E, elementary). While this scale has been repeatedly used and adapted, it is important to note all versions have been developed under a simplistic, one-dimensional conceptualization of mathematics anxiety.

Wigfield and Meece (1988), Mathematics Anxiety Questionnaire (MAQ)

Recognizing the problems of relying on a one-dimensional conceptualization of mathematics anxiety, Wigfield and Meece (1988) looked to build off of Liebert and Morris' (1967) work on the affective and cognitive dimensions of test anxiety. In doing so, they ultimately developed the 11-item Math Anxiety Questionnaire (MAQ) containing seven items that tap the affective dimension of mathematics anxiety and four items that tap the cognitive dimension. In their study of 6th and 12th grade students, Wigfield and Meece (1988) used exploratory factor analysis (EFA) to find two unique factors and then confirmatory factor analysis (CFA) to confirm the two-factor model. Using CFA, they also found invariance of the model across age group (6th versus 12th grade) and gender (boys versus girls). The reliability alphas for the affective and cognitive subscales were 0.82 and 0.76, respectively. While Wigfield and Meece (1988) looked to extend thinking about the dimensionality of the mathematics anxiety construct, they did so within the classical test theory framework of instrument development. Next, we look at more recent scales developed with Rasch measurement theory, and the Rasch rating scale model in particular.

Prieto and Delgado (2007), Math Anxiety Scale (MAS, *Test de Ansiedad hacia las Matemáticas*)

Prieto and Delgado's (2007) Math Anxiety Scale (MAS) contains 18 items that describe various behaviors, beliefs, and psycho-physiological reactions related to both academic and daily use of mathematics. The scale was developed in Spanish and administered as part of two studies in Spain, first to a group of mostly 13- and 14-year-olds and second to a group of mostly 15-year-olds. The first administration of the scale used a 6-point Likert scale from *absolutely disagree* to *absolutely agree*; in the second administration, the response options were reduced to a 4-point scale (with the same anchors). From Study 2 (with the revised response options), the MAS showed acceptable indices of reliability called item and person separation reliability. The MAS also showed invariance across achievement and gender with the exception of two items. Removing these two items, the scale still showed good psychometric properties. Specifically, "item and person separation reliability are adequate (.99 and .88, respectively), and anxiety-achievement and anxiety-gender correlations remain invariant (-.27 and -.23, respectively)" (Prieto & Delgado, 2007, p. 155). While the instrument was analyzed with the Rasch rating scale model, the instrument itself, at least as described by the authors, was not developed using the ideas and principles Rasch measurement theory, which will be described later.

Cavanagh and Sparrow (2011)

So far we have noted three limitations in previously developed instruments: 1) application of simple one-dimensional conceptualizations of mathematics anxiety, 2) reliance on classical test theory methods in data analysis, and 3) failure to use Rasch measurement theory in instrument development even when Rasch analysis is employed. Cavanagh and Sparrow (2011) show the first clear indication of mitigating all three of these existing limitations. While they do not explicitly name Rasch measurement theory in their development of the construct model, Cavanagh and Sparrow (2011) employ some of the theory's principles, which will be discussed later, such as variation in item difficulty and hierarchical in the nature of their progression along the continuum (which they call the developmental assumption). In their construct model of mathematics anxiety, they theorize mathematics anxiety to occur within three broad situations: instruction, assessment, and application (outside the classroom). Within each of these situations, individuals can experience various levels of mathematics anxiety, from low anxiety, shown by attitudinal indicators, to moderate or high anxiety, with cognitive indicators, up to very high anxiety, exhibited by somatic indicators. Note that these levels (attitudinal, cognitive, and somatic) are types of indicators that we have seen before although conceptualized as separate components rather than levels of a single one-dimensional mathematics anxiety construct. They theorize that these levels are cumulative, although lower levels may be overshadowed by higher level indicators and therefore not visible.

After defining the construct model, Cavanagh and Sparrow (2011) developed 21 items: six attitudinal, nine cognitive, and six somatic. More cognitive items were developed as these items are theorized to be at the moderate level of mathematics anxiety and would make the instrument more sensitive to students with average or moderate levels of mathematics anxiety. Individuals are given four response options from *strongly agree* to *strongly disagree*. Based on their sample of 50 children from seven primary schools in Perth and using Rasch analysis, eight items were removed because of poor fit to Rasch analysis criteria, which will be presented later, resulting in thirteen items in the final analysis. In addition, based on these analyses, Cavanagh and Sparrow (2011) found attitudinal, cognitive, and somatic indicators at all levels of mathematics anxiety. Therefore, their previous assumption of a cumulative relationship between the attitudinal, cognitive, and somatic indicators was rejected in favor of a multidimensional construct model where the three domains of indicators are seen as separate dimensions and the levels of anxiety are simply low, moderate, and high anxiety.

While much has been done on the measurement of mathematics anxiety, additional work is needed to develop and confirm a mathematics anxiety construct model and scale. Thus far, most instruments to measure mathematics anxiety have been developed without a strict definition of the construct and using classical test theory methods. Rasch measurement theory and the Rasch rating scale model show promise for the development of a theoretically grounded scale of mathematics anxiety. Rasch measurement theory and its advantages over classical test theory are discussed in the next section.

Rasch Measurement Theory

Rasch measurement theory was developed by Georg Rasch (1960/1980) and has been expanded upon and clarified by others (e.g., Andrich, 1978; Ludlow et al., 2014; Wright & Masters, 1982). According to Wright and Masters (1982), "measurement begins with the idea of a variable or line along which objects can be positioned, and the intention to mark off this line in equal units so that distances between points on the line can be compared" (p. 1). This idea has been translated into a "variable map" where the line defines the continuum of the construct being measured and the objects positioned along that line are items. The goal is to write items that span the entirety of the construct and therefore spread along this line to describe and compare individuals measured with the scale, or items. In order to objectively compare individuals and usefully predict and diagnose, the construction of the variable map needs to be systematic, and the variable map needs to be reproducible, such that the person measures and item difficulties are the same, across persons and time points (Wright & Masters, 1982). Rasch measurement theory was developed with the goal of objective measurement at the forefront (Wright, 1980). Objective measurement in this context means that an instrument to measure a latent trait provides unbiased depictions of individuals' current state of the latent trait for purposes of diagnosis, comparison, and/or prediction (Wight & Masters, 1982).

Separability and Sufficiency

Under Rasch measurement theory, person ability estimates must not depend on the difficulty of the items, and item difficulty estimates must not depend on the individuals responding to the items (Wright & Masters, 1982). This phenomenon is known as separability and is achievable with Rasch measurement models, including the Rasch rating scale model (Wright & Masters, 1982). Because the Rasch rating scale model is a probabilistic model where person and item parameters are additive, the estimation of person and item parameters can be separated from one another (Rasch, 1966). In this way, the total raw scores of the items are sufficient for obtaining item parameter estimates and the total raw scores of the persons are sufficient for obtaining person parameter estimates; this is known as sufficiency (Rasch, 1966). Objective measurement can be achieved only with separability and sufficiency, as these characteristics result in measurement invariance through "sample-free test calibration" and "test-free person measurement" (Wright & Masters, 1982). The seven Rasch measurement principles that aid in the development of objective measures are discussed below.

Rasch Measurement Principles

The principles of Rasch measurement for optimal item development in the generation of an instrument are: 1) unidimensionality, 2) variation, 3) uniform spread of the items along a continuum, 4) hierarchical in the nature of their progression along the continuum, 5) equally discriminating, 6) independent, and 7) well-fitting as a match between theory and data (Rasch, 1960/1980; Ludlow et al., 2014, p. 129). Unidimensionality means that only one construct, or variable, is being measured. Specifically, under Rasch assumptions, the construct of interest is not being mixed with other constructs. The second principle, variation, indicates that the items should vary in their difficulty – for example, from very easy to very hard items. Additionally, items should have a uniform spread. This means that the easy and hard items are spread evenly across the continuum of the construct. Fourth, the items should be hierarchical in their construction and should progress from easy to hard. Next, the items should be equally discriminating. If variation in item discrimination is found, a second parameter (a discrimination parameter) could be added to the model creating a 2-parameter model, but it would then no longer be a Rasch 1-parameter model. The sixth principle is local independence of the items, which means that the response on one item should not depend upon or affect the response on any other item. Finally, one should ensure that the items, or data, fit well to the model according to the theory; that is, the empirical variable map should correspond closely to the conceptual variable map. If there is mismatch between the data and the theory, then the items should be analyzed to see how they could be revised to better fit the theory, or the theory may have to be revisited. Together, these principles establish how an instrument should be constructed according to Rasch measurement theory.

Advantages of Rasch Measurement Theory over Classical Test Theory

A brief introduction to classical test theory. Classical test theory (CTT), or true score theory, is based on the concept that one's observed score (X) is made up of two unobservable components: one's true score on the latent construct of interest (T) and measurement error (E), where the relationship between these components can be represented as $X = T + E$ (Lord & Novick, 1968). An assumption of CTT is that an individual's latent trait of interest remains unchanged (that is, T is constant), while the observed score (X) from a single test administration is a chance variable with an unknown frequency distribution (Lord & Novick, 1968). The expected value of this frequency distribution, or the expected value of the observed scores, is one's true score (T). Based on the linear equation, the measurement error (E) is the difference between one's observed score and true score. The goal of CTT is to reduce measurement error so that the observed score is as close of an approximation to the true score as possible.

Under CTT, items on a test or scale are designed to be parallel measurements of the same thing. That is, they are intended to be replications of the construct. Therefore, items should be highly correlated with each other and relatively moderate in their difficulty (Lord & Novick, 1986). In addition, CTT relies on statistical methods that are based on overall test or scale scores; that is, item estimates, person estimates, and reliability estimates are based on test-level statistics (i.e., composite scores) and utilize raw score calculations (e.g., mean, standard deviation, and variance) (Hambleton & Jones, 1993). Because CTT relies on methods that utilize raw scores for item and person estimates, these estimates are sample-dependent and test-dependent, respectively. That is, the ability level of the group will directly influence the item estimates, and the difficulty of the assessment or scale will directly influence the person estimates. Therefore, for a test to give reliable information about the respondents or for the sample to give reliable information about the test, there must be a match between the difficulty of the items on the test and the ability level of the persons in the sample. Any mismatch would affect the reliability of the estimates and the usefulness of the information given. Furthermore, the test-dependency of the person scores limit their usefulness for comparing different groups or looking at individual change over time (Hambleton & Jones, 1993), and the sample-dependency of the item and test statistics limit their utility for revising specific items or the overall test.

Comparing Rasch measurement theory and CTT. One of the core differences between Rasch measurement theory and CTT is how constructs, and therefore items, are defined. In CTT, items are intended to be replications of the construct such that all items are highly correlated and have a moderate level of difficulty (Lord & Novick, 1968). In contrast, Rasch measurement theory views constructs along a hierarchical continuum for which items should be created to capture the various levels of construct difficulty (Wright & Masters, 1982). CTT relies on correlations between items to determine adequacy of the instrument; whereas Rasch measurement theory looks more specifically to see whether item difficulties order according to the hypothesized theory. The variable map available with Rasch analysis provides a stronger check of the construct validity than CTT measures. This direct test of the fit between the hypothesized model and the observed data is an important advantage of methods associated with Rasch measurement theory.

Another main difference between CTT and Rasch measurement theory is in how item and person estimates are determined. In CTT, item and person estimates are raw scores of responses (for example, item estimates are often the mean of person responses to the item and person estimates are the raw sum score of an individual's responses to all items). In contrast, Rasch measurement theory relies on probabilistic models which allow for the log transformation of raw scores into *logits*, or logistic units, which is an interval measurement scale (Wright & Masters, 1982). Logits are linear such that a one-logit difference between persons or items regardless of where it is at on the scale has the same probability of an item being answered correctly. This is in contrast to raw scores in CTT which do not have the property of linearity, and therefore, a one-point difference can have a vastly different probability of a correct response depending on which part of the score distribution is being looked at (Wright & Masters, 1982).

Related to this, because Rasch measurement theory employs probabilistic models and CTT does not, Rasch measurement theory has the additional advantage of being able to calculate expected scores for a person on an item and comparing that to the observed score, which can then be used to assess the fit of the person and item statistics. This makes it easier to detect problems with the instrument and make item-specific revisions as needed. Analyses used to assess instruments developed using CTT do not have the ability to assess fit of individual persons and items using observed and expected scores, and therefore are not as useful for instrument revision.

A final advantage of Rasch measurement theory over CTT is in the fact that item and person estimates under Rasch measurement theory are person-free and test-free (Wright & Masters, 1982). Therefore, information about individual persons and items is applicable outside of the specific context of the instrument and sample. In contrast, item and person estimates in CTT are sample dependent and test dependent. Sample independence is particularly important when developing an instrument. Because item estimates in CTT are sample dependent, trying to make changes to items can be like trying to hit a moving target, as each new sample could provide different results. In contrast, item estimates under Rasch measurement theory are sample independent and therefore changes to an item will hold across samples, making item revisions easier. Sample and test dependence is particularly problematic for CTT when trying to compare performance across groups or measure change over time. That is, an instrument that is not 'fixed' and is sample dependent is not useful as the comparisons then become relative to the context. In contrast, because person and item estimates are test-free and sample-free and have sufficient statistics for person and item parameters, scores can be used to objectively compare different groups or performance on an instrument over time. This is perhaps the greatest advantage of instruments developed with Rasch measurement theory instead of classical test theory.

Initial Instrument Development for the Comparative Math Anxiety Scale

Developing the Construct

To develop a reliable and construct valid Rasch-based scale, one must understand the construct in enough depth to create a “construct map” (Wilson, 2005). A concept, or construct, map is an organizational tool that makes tangible the idea (concept or construct) of interest through definition and description of its various components and levels. The task of developing a construct map for mathematics anxiety was undertaken by another doctoral student and me as part of a sequence of psychometrics courses. We achieved this aim by reviewing the previously described literature on mathematics anxiety as well as relying on previous teaching experience, experience as students in mathematics courses, and our own understanding of the construct of anxiety generally and specifically as it applies to the field of mathematics. As described previously, Cavanagh and Sparrow (2011) conceptualized mathematics anxiety to have three levels: physical or somatic indicators of anxiety representing extreme anxiety (the highest level), cognitive indicators of anxiety or anxious mental processes representing high anxiety (the middle level), and attitudinal or emotional indicators of anxiety representing low anxiety (the lowest level). Cavanagh and Sparrow (2011) theorized that these levels were cumulative such that a person exhibiting indicators of a higher level of anxiety would also display indications of the lower level(s) of anxiety; however, the lower level(s) could also be overshadowed by indicators of the higher level(s) of mathematics anxiety and therefore not be visible. They later revised their conceptualization of mathematics anxiety to include somatic, cognitive, and attitudinal mathematics anxiety as separate dimensions of mathematics anxiety that all span from low to high anxiety. Dr. Caroline Vuilleumier and I used this work as a basis for the development of our own mathematics anxiety framework.

In conceptualizing mathematics anxiety for the development of our instrument, the decision was made to think of physical, mental, and emotional anxiety not as levels but as distinct categories, or dimensions, of mathematics anxiety that *all* can range from extremely high to extremely low, or non-existent anxiety. Here, the three dimensions of mathematics anxiety are distinct but highly correlated. Five levels were theorized to occur across and within each of the three dimensions. From high to low anxiety, these levels are: overwrought, distressed, antsy, apprehensive, and tranquil (see *Table 1* on page 8 for the mathematics anxiety construct maps for the three dimensions).

Based on our conceptualization, an individual at the highest level of mathematics anxiety is overwrought, and shows high levels of emotional, mental, and physical mathematics anxiety. This means that the individual feels emotionally overwhelmed, cannot recall what one learned as one's mind keeps blanking out, and is on the verge of tears or is physically ill. In contrast, an individual at the lowest level of mathematics anxiety is tranquil, and exhibits no (internal or external) symptoms of anxiety. Specifically, individuals with low mathematics anxiety have no emotional feelings of anxiety, can think clearly and remain focused, and exhibit no physical symptoms of anxiety when confronted with or when doing math problems. Please note that the construct map and continuum was initially created with the development for a measure of mathematics testing anxiety in mind; however, the construct map proposed is not specific to testing scenarios and can be applied without revision to the development of a general mathematics anxiety scale.

Mapping Items to the Continuum

Using the construct map as a starting point, items were written to capture each level of mathematics anxiety across all three dimensions. Initially, the language of the level description was used. When that became insufficient, a thesaurus was used to generate additional items. In addition, we relied on the mathematics anxiety literature, our own knowledge of anxiety generally and mathematics anxiety specifically in our experiences from obtaining a combined mathematics and psychology degree, and in the other graduate student's experience as a high school mathematics teacher. Using this method, twenty-five items were developed: eight measuring emotional/attitudinal math anxiety, nine measuring mental/cognitive math anxiety, and eight measuring physical/somatic math anxiety (see *Table 2* on page 9 for a table of the mathematics testing anxiety items organized by dimension and level). Note, in *Table 2*, that if more than one item exists in a level for a dimension then the items are organized by the order in which they appeared on the survey and not by some theorized level of difficulty. This is either because the two items are theorized to be at the same level or a specific expectation about their ordering has not been determined. More items were developed to measure the middle three levels of the scale than the highest and lowest levels of the scale, as more individuals are expected to fall in this moderate range and therefore it was expected that more items would be needed to distinguish between these moderate levels of math testing anxiety.

Typically, items for most instruments are written from the first-person perspective ("I") and have corresponding Likert response options of *strongly agree* to *strongly disagree*. However, these item types are susceptible to response biases, most notably social desirability bias and acquiescent responding (Paulhus, 1991). Social desirability bias refers to the tendency to respond in a manner that is perceived as desirable by others (Kuncel & Tellegen, 2009), while acquiescent responding refers to the preference for responding, regardless of item content, on the positive side of the rating scale (Weijters et al., 2013). In our instrument development process, we chose to design items as third-person narratives, where respondents are asked to compare themselves to individual ‘X’. This decision was made with the expectation that these changes would avoid some of the above challenges often seen with traditional Likert-type scales.

To go along with the third-person narratives, specific instructions with unique, *comparative* response options were required. Adapting from Ludlow et al. (2014), survey-takers saw the following instructions: Each question will provide a scenario describing how a person feels, thinks, or behaves while he/she is taking a mathematics exam. You will be asked: On a typical exam day, are you (1) much less anxious than ‘X’, (2) less anxious than ‘X’, (3) as anxious as ‘X’, (4) more anxious than ‘X’, or (5) much more anxious than ‘X’. An example item from the distressed level of the mental/cognitive dimension is “Joe feels like he should know how to solve the problem but he can’t put his thoughts in order.” Response options for this item are: much less anxious than Joe, less anxious than Joe, as anxious as Joe, more anxious than Joe, and much more anxious than Joe.

Results of Initial Item Development

The items developed went through two rounds of data collection in fall 2013 and fall 2014 with undergraduate students enrolled in a mathematics course at a private 4-year institution in the United States. There were 49 and 66 students in the first and second samples, respectively. Results from these data collection efforts were presented at the American Educational Research Association (see Vuilleumier & Klein, 2015 for a full presentation and discussion of the results). Given the small sample sizes from our two data collection periods, we looked into combining the samples for data analysis. However, inspection of the variable maps (see *Figures* 1 and 2 on the following pages) showed some variation and disordering of items across the two samples (bolded items are disordered across the two samples); therefore, it was deemed inappropriate to combine the samples and instead two sets of analyses were carried out for the scale.

Figure 1

Comparative Mathematics Testing Anxiety Variable Map (Sample 1)

```
MEASURE    Student - MAP - Item
                <more>|<rare>
                  30  +
                      |
                      |
                      |
                      |
              41   8 T|
 5                    +                                   A3_2                      Overwrought
                      |T                                  A3_6
                      |
                  21  |
                      |
              44  51  |                                   A3_5
 4            15  23  +                                   A3_7                       Distressed
                     S|
          32  38  45  |                 A2_6              A3_1  A3_8
                      |
              13  49  |                 A2_1
              10  39  |
          14  24   5  +M A1_1  A1_4     A2_4
                  19  |  A1_6  A1_8     A2_5  A2_8        A3_3
          11  28  42  |  A1_5           A2_9
          16  35   4 M|  A1_2  A1_3
              33  34  |
 3            18   9  |S                A2_3                                            Antsy
              36  48  +                 A2_2  A2_7
                  17  |
       1  12  25  37  |                                   A3_4
              20  26  |
              43   6  |  A1_7
               2   7 S|T
                      +
       27 29  40  50  |
 2                31  |                                                         Apprehensive
                      |
                      |
                   3  |
                      +
                  22 T|
                      |
                      |
                      |
                      |
                      +                                                             Tranquil
                <less>|<frequent>
```

Figure 2

Comparative Mathematics Testing Anxiety Variable Map (Sample 2)

```
MEASURE                                      PERSON - MAP - ITEM
                                                 <more>|<rare>
    3                                             37   +
                                                       |                                A3_2
          Overwrought                                  |
-------------------------------------------------------------------------------------------------
                                                       |
                                                       |T
    2                                                  +
                                                       |
                                                       |                                A3_6
          Distressed                              64   |                                A3_5
-------------------------------------------------------------------------------------------------
                                                       |
    1                                                T+S                                A3_1
                                                       |                A2_6            A3_7
                                                  26   |                A2_9            A3_8
                                   2   6  25  47  58   |                A2_1  A2_4
                                                       |  A1_1          A2_8
    0                                         30  65 S+M  A1_4          A2_5
                                          13  46  63   |
                                      10  16  50  56   |  A1_5  A1_8                    A3_3
          Antsy               11  17  33  36  43  54   |  A1_2  A1_3    A2_3
-------------------------------------------------------------------------------------------------
                               8   9  14  18  32  68   |  A1_6
   -1                                 40  57  61  62 M+S
                           5   7  12  15  35  42  48   |                A2_7
                                      22  23  38  39   |
       4  19  24  27  29  31  41  44  49  51  53  60   |                A2_2            A3_4
                                                  20   |
   -2                                                 S+  A1_7
                                              21  55   |T
                                       1   3  34  59   |
          Apprehensive                            28   |
-------------------------------------------------------------------------------------------------
                                                  52   |
   -3                                                 T+
                                                       |
                                                       |
          Tranquil                                     |
                                                  45   |
   -4                                                  +
                                                 <less>|<frequent>
```

After looking at the variable maps to determine if it was appropriate to combine the samples, the variable maps were then used to look at the convergence of the results with the *a priori* theory of the construct definition. Overall, the variable map did not match the theoretical construct map. Specifically, it was theorized that all three dimensions (physical, mental, and emotional) would span all five levels of the mathematics anxiety construct; however, it instead appeared that while all three dimensions did span multiple levels, they were also ordered in their leveling. That is, items measuring physical mathematics anxiety (A3_#) spanned the highest levels, items measuring mental mathematics anxiety (A2_#) spanned the middle levels, and items measuring emotional mathematics anxiety (A1_#) also spanned the middle levels but with its hardest item (A1_1) being quite a bit easier than the hardest mental mathematics anxiety item (A2_6) in both samples. This suggests a potential adjustment to the theory, such that the three dimensions are leveled, as Cavanagh and Sparrow (2011) originally suggested, but with some overlap. In addition, no items captured the lowest level of mathematics anxiety; therefore, additional work is needed to capture very low student anxiety.

The variable maps were also used to look at disordering of the items (high anxiety = overwrought = yellow, distressed = green, antsy = blue, apprehensive = purple, tranquil = grey = low anxiety). While it has been already noted that the items were disordered across all the dimensions, I have yet to note that the items were also disordered within each dimension. For example, in the emotional/attitudinal mathematics anxiety dimension, item A1_6 was theorized to be in the highest level; however, in sample 1, it was less difficult than items A1_1 and A1_4 and as difficult as item A1_8, and in sample 2, it was the second easiest item only having a higher difficulty than item A1_7 (the theorized easiest item). Similar disordering occurred within both the mental/cognitive and physical/somatic mathematics anxiety dimensions. Together with the previous finding of leveled dimensions, this suggests serious revision is needed to the theory and to the items used to measure the construct. It is important to keep in mind that these samples were obtained from undergraduate students so there is a possibility that the items would behave differently for pre-college students. Therefore, additional data collection may be required before major changes are made to the conceptual framework.

Summary

In the first chapter I presented some of the challenges and limitations of defining and measuring mathematics anxiety. This chapter reviewed the relevant areas in the literature that justify the need for a Rasch-based scale that measures high school students' anxiety in mathematics, including a review of mathematics anxiety as it has been conceptualized, as it relates to other constructs, and as it has been measured previously. This chapter concluded with an argument for Rasch measurement theory and methods, as well as a description of the initial instrument development work that lays the groundwork for this dissertation. Given the current state of the work to measure mathematics anxiety, the overarching research question is: To what extent can the construct of mathematics anxiety be measured reliably and meaningfully by developing a Rasch-based scale? In the methodology chapter, I describe the overall research design; the participants including the population of interest, the sampling approach, and the final study samples; the procedures for instrument development, data collection, and data analysis; and end with a discussion of quality assurance and ethical considerations.

CHAPTER 3 – METHODOLOGY

This chapter begins with a re-statement of the research questions followed by a description of the overall research design and the larger research study within which this dissertation took place, as well as an examination of the target population and sampling approach. Given the extensive process of instrument development, an entire section is dedicated to the procedures that were followed for instrument revisions and potential changes to the conceptual framework of mathematics anxiety. This section includes a description of the item revision process. Then, I describe the data collected, both quantitative (secondary survey data) and qualitative (primary cognitive interviews). Next, I describe the analyses conducted to answer the research questions. Finally, I end with a discussion of data quality and research ethics.

Research Questions

Again, the broad research question that guided this dissertation is: To what extent can the construct of mathematics anxiety be measured reliably and meaningfully by developing a Rasch-based scale? Specific research sub-questions are:

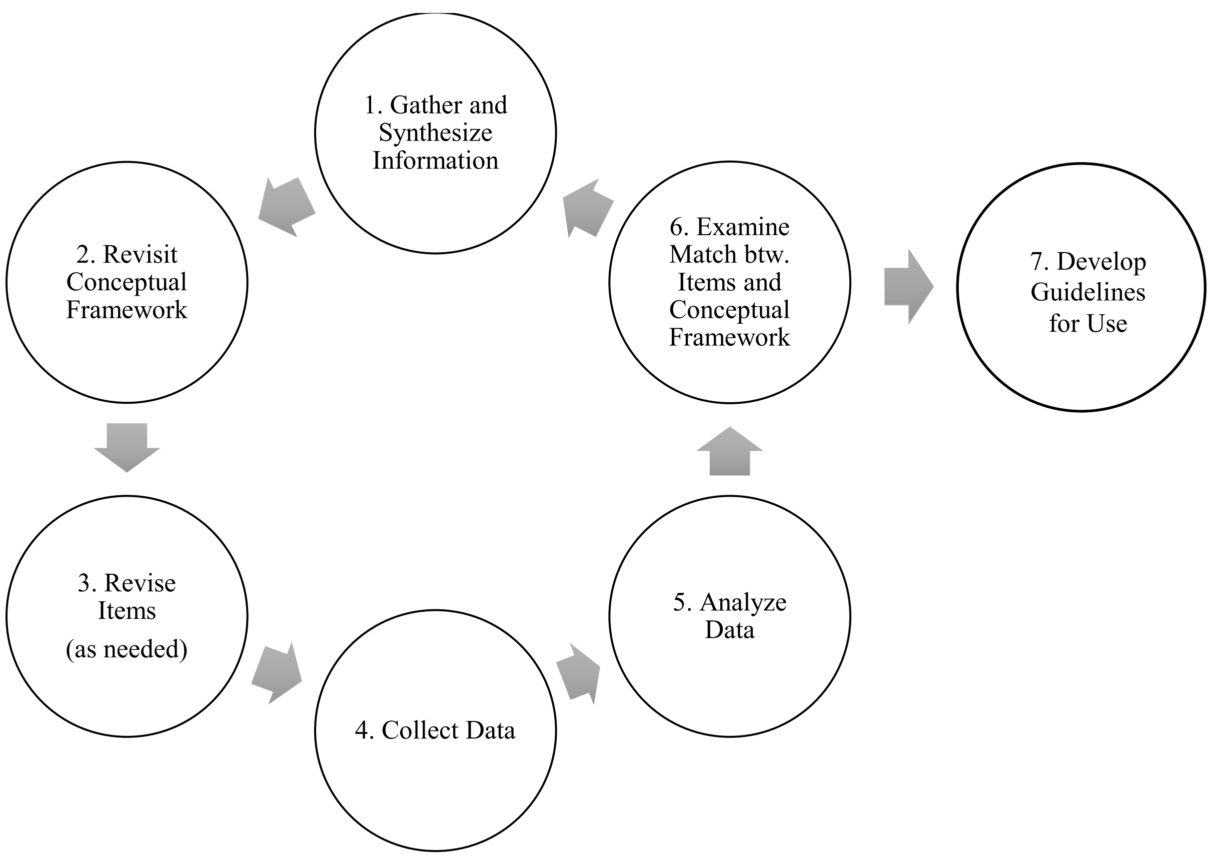
1. How can the tripartite conceptual framework of mathematics anxiety Dr. Caroline Vuilleumier and I developed be employed and adapted in the development of a Rasch-based scale;
2. Is the scale invariant across grade, gender, and other student demographics; and
3. Does the scale have an acceptable degree of construct and criterion validity?

Research Design

The purpose of this dissertation is to develop and validate an instrument to measure students' mathematics anxiety. This requires three phases: 1) initial construct definition and item development; 2) construct clarification and item revision; and 3) instrument validation. Both the first and second phases of the dissertation follow the iterative scale development process shown in *Figure 3* below, which was adapted from the process used in Sondergeld and Johnson (2014). Through this process, information about the construct of interest is gathered and synthesized (Step 1). This information comes from expert knowledge, review of the literature, and existing data. Next, the conceptual framework is revisited (Step 2) before items are revised if needed (Step 3). Finally, pilot data are collected (Step 4) and analyzed (Step 5) to allow for the examination of the match between the items and the conceptual framework (Step 6). After Step 6, we either determine that more work is needed and cycle back to Step 1 or determine that there is adequate match between the items and the conceptual framework and provide guidelines for how the scale should be used (Step 7).

Figure 3

Iterative Scale Development Process



The first phase, initial construct definition and item development, was completed prior to the dissertation, as described in the literature review chapter. This phase followed the iterative scale development process described above. The second phase, construct clarification and item revision, began during the proposal stage of the dissertation, and spanned several years (see *Table 3* below for the complete scale development timeline). A more complete description of the iterative scale development process is presented as a separate section later in this chapter. The third and final phase of the study is instrument validation. After the final round of item revisions, the Comparative Mathematics Anxiety Scale was administered along with several validation scales. Assessment of the match of the data to the conceptual framework and the guidelines for scale use serve as the evidence for the construct validity of the scale, while the validation scales provide evidence for the criterion validity of the measure.

Table 3

Study Timeline

Scale Development Process		Revision Timeline			
Step	*Description*	*Round 1*	*Round 2 (survey)*	*Round 2 (interviews)*	*Round 3*
1	Gather and synthesize information	W 2018	SU 2018	F 2019	SU 2020
2	Revisit conceptual framework	W 2018	SU 2018	n/a	SU 2020
3	Revise items (as needed)	SP 2018	SU 2018	n/a	F 2020
4	Collect data	SP 2018	F 2018	F 2019	F 2020
5	Analyze data	SU 2018	W 2019	W 2020	W 2021
6	Examine match between items and conceptual framework	SU 2018	SP and SU 2019	SP 2020	W 2021
7	Develop guidelines for use	n/a	n/a	n/a	SP 2021

W = winter; SP = spring; SU = summer; F = fall

Supporting Success in Algebra **Study**

In this dissertation, the Comparative Mathematics Anxiety Scale was developed and piloted as part of a four-year research project funded by the NSF [NSF 1621011]. The NSF project, *Supporting Success in Algebra*, aims to examine the impact of a year-long algebra support curriculum, Transition to Algebra (TTA), on 9th-grade students' algebra performance and their attitudes towards mathematics (Mark et al., 2014). Specifically, this project is interested in the following student attitudes towards mathematics: anxiety, confidence or self-concept, perseverance, internal and external motivation, growth mindset, and beliefs about mathematics. TTA is designed to increase student performance in algebra, as well as improve students' attitudes toward mathematics.

Students' algebra performance and attitudes toward mathematics will be compared across the TTA and business-as-usual groups in a quasi-experimental, pre-post research design, building off of prior work that utilized a regression discontinuity design (Louie et al., 2016). The first year of the study (2016-2017) was devoted to planning the implementation of the study; the second year of the study (2017-2018) to instrumentation piloting; the third year (2018-2019) to full-scale data collection; and the fourth year (2019-2020) to additional support, as well as data analysis and dissemination that has continued into a fifth year (2020-2021). The *Supporting Success in Algebra* project served as the ideal platform for the development, testing, and revision of the Comparative Mathematics Anxiety Scale, providing the opportunity for multiple administrations that were already occurring as part of the NSF study.

Participants

Target Population and Sampling

The target population of this dissertation is all high school students in the United States. This dissertation study took place alongside the larger NSF-funded *Supporting Success in Algebra* study, which includes 9^{th}-grade students in districts across the United States. Data was collected from this sample for one round of revisions, with the other data collection rounds occurring with pilot districts in the greater-Boston area. The sample size obtained from this partnership must be large enough to provide reliable and valid estimates of item statistics. The suggested sample size for instrument development is not concrete. However, Crocker and Algina (1986) suggest a minimum of 200 for an item analysis study, while Nunnally (1967) provides the general rule of thumb to be five to ten times the number of subjects as items, with fifteen to twenty times being ideal (Clark & Watson, 1995; DeVellis, 2003; Hair et al., 2010; Morgado et al., 2017). The number of items for each of the scales varies across administrations from seven to thirteen items (this will be described in detail in the next chapter). Therefore, a sample size of 130 high school students would be acceptable, with 200-300 students being ideal.

The Study Samples

Round 1 sample. For the first pilot of the Comparative Mathematics Anxiety Scale in spring 2018, 77 high school students were recruited from a local district in the greater-Boston area to complete a survey aimed at developing the *Opinions about Mathematics Survey* for the *Supporting Success in Algebra* study. Thirty-seven of the students identified as female (48.1%) and thirty-eight students identified as male (49.4%), with gender data not provided for two students.

Round 2 samples. For the second pilot of instrument revisions in fall 2018, 1,588 total 9th-grade students completed the *Opinions about Mathematics Survey* as part of the *Supporting Success in Algebra* study in districts across the country. Each dimension of the Comparative Mathematics Anxiety Scale was administered on a different form of the survey—596 students completed the emotional-attitudinal subscale, 845 students completed the mental-cognitive scale, and 147 students completed the physical-somatic scale. Also as part of this round of data collection, fifteen students participated in cognitive interviews in fall 2019 at a local district in the greater-Boston area—five students by completing one of the three forms of the fall 2018

Opinions about Mathematics Survey then participating in an interview and ten students by participating in an item-sorting activity while being interviewed.

Round 3 sample. The third and final administration of the Comparative Mathematics Anxiety Scale took place in fall 2020 at a local district in the greater-Boston area. One-hundred and thirty-two (132) high school students participated in the study and completed at least one of the three subscales in the survey. The grade breakdown of the sample is as follows: 29 freshman (22.0%), 16 sophomores (12.1%), 53 juniors (40.2%), and 27 seniors (20.5%). Sixty-eight (51.5%) students identified as female, 55 (or 41.7%) as male, and two students (1.5%) preferred not to answer. For English-language proficiency (ELP), 113 (85.6%) students were categorized as English-language proficient and 12 (9.1%) students were categorized as not English-language proficient. Seven students (5.3%) did not provide demographic information.

Iterative Scale Development Process

The scale development process used here extends the process utilized in Sondergeld and Johnson (2014) by making the process iterative, such that the match between the items and the conceptual framework is checked (now Step 6), and if the match is not adequate, the process goes back to Steps 1-6. Only after the match between the items and the conceptual framework has been deemed adequate, does the scale develop process proceed to the seventh and final step—develop guidelines for use. Each of the seven steps that make up the iterative scale develop process for this dissertation are described in more detail below.

Step 1: Gather and Synthesize Information

The first step in developing a reliable and valid measure is to understand as fully as possible the concept of interest, in this case, mathematics anxiety. The initial step of surveying the literature was carried out during the initial instrument development work described in the literature review chapter. For later revisions to the instrument developed through this work, step one will involve review of and reflection on the literature, discussion with content knowledge experts, and review of the previously collected data (survey and/or interview data).

Step 2: Revisit Conceptual Framework

The second step is to create, reflect on, and/or revise the conceptual framework for the construct. In the first round of instrument development, the conceptual framework is either created, borrowed, or adapted from another source. In this case, Dr. Caroline Vuilleumier and I developed a conceptual framework based on the suggestions Cavanagh and Sparrow (2011) provided through their own instrument development work. In subsequent iterations through the scale development process, it will be important to revisit this conceptual framework. Not every iteration will result in revisions to the conceptual framework; however, if the evidence from the data collected supports the need for adjustment, which may or may not also be supported by evidence in the literature, edits to the conceptual framework should be made at this step.

Based on the existing results of initial instrument development, the scale revision, administration, and interviews to be conducted, and reflecting on theories of anxiety generally and mathematics anxiety specifically, mathematics anxiety, as conceptualized here, will likely be re-conceptualized to fit more in line with Cavanagh and Sparrow's (2011) re-conceptualization. That is, the five theorized levels (from high to low: overwrought, distressed, antsy, apprehensive, and tranquil) of the three dimensions of mathematics anxiety (physical, mental, and emotional) may be scaffolded from the original conceptualization of mathematics anxiety (see *Table 1* on page 8) to take into account the leveling of the dimensions that was found in the initial instrument development work, and that will likely be confirmed in the first round of the instrument revision work. That is, the highest level of anxiety would be defined solely by physical/somatic indicators, while the lowest level of anxiety would be represented solely by emotional/affective indicators, where a high score still indicates high mathematics anxiety. Specifically, instead of spanning all five levels, the intention is that the physical indicators would span the highest three levels (overwrought, distressed, and antsy); the mental indicators would span the middle three levels (distressed, antsy, and apprehensive); and the emotional indicators would span the lowest three levels (antsy, apprehensive, and tranquil).

Step 3: Revise Items (as needed)

Once the available data, literature, knowledge from content experts, and conceptual framework have been reviewed and considered, it is important to next consider revisions to the items as the third step in the iterative scale development process. Using the prior steps as a guide, changes should be made to the items if the scale does not yet align to the conceptual framework. Item revisions may involve writing new items to fill in gaps, removing existing items that either do not fit with the conceptualization or are repetitive with other existing items, or editing existing items for clarity or to fill in gaps. Item revisions will depend on the needs of the scale and the results of the prior round of data collection. As needed, new items will be generated by looking for synonyms to existing items or phrases in the conceptual framework. Additionally, if items are revised for clarity or to fill gaps this will occur by clarifying the language or adding adjectives to the sentence to make the item more or less difficult. Appropriate adjectives might be completely or extremely (to make the item more difficult) and slightly, a little, or a bit (to make the item easier). Item writer judgment is always a consideration in the item revision process; that is, it is unlikely that two item writers would make the exact same adjustments to the scale and to each item more narrowly.

Step 4: Collect Data

Once steps one, two, and three have been completed, data are collected to gather evidence about the appropriateness and success of any changes that were made. Data was collected in one of two forms: survey responses or cognitive interview. Survey data is useful for examining the ordering of the items and whether they match the ordering hypothesized based on the conceptual framework of mathematics anxiety. However, sometimes survey data does not provide enough information about the reasons for why students respond in a particular way; this is especially the case when items do not order as expected. In this situation, it will be important to conduct cognitive interviews of the survey items. The purpose of these interviews is described in the data collection section, but generally the purpose is to further understand how high student students understand the construct of mathematics anxiety and the items being presented to them in the mathematics anxiety scale. Interviews serve as an additional method to aid in the development of and revisions to the scale as well as the underlying conceptualization of the construct of mathematics anxiety.

Step 5: Analyze Data

Once the data had been collected, it was analyzed according to the procedures laid out in the data analysis section presented later in this chapter.

Step 6: Examine Match between Items and Conceptual Framework

The final step in the iterative scale development process is to examine the data collected and analyzed to assess the match between the items and the conceptual framework. Under Rasch measurement principles and Rasch rating scale analysis, this is achieved by examining a variable map of the data collected (Ludlow et al., 2014, 2019). The variable map provides a means for looking at whether the scale is appropriate for the population of interest and whether the items of the scale match the conceptualization of the construct. First, one should determine that the item distribution matches the person distribution. That is, the scale as a whole should not be too hard or too easy for the population of interest. Next, the items should order according to the conceptual framework. If these do not match, changes to the conceptualization of the scale, changes to the conceptualization of the construct, or changes to both should be considered, as illustrated in Ludlow et al. (2014, 2019) and Matz-Costa et al. (2014). The framework used in this dissertation is new and therefore should be considered in-progress and open to revision given appropriate evidence. Finally, the items should spread evenly across the distribution of people. If there are gaps between items or clumping of items, additional revisions to the items should be considered. Based on how well the items match the conceptual framework, one should go back to step one or move on to collecting evidence for other aspects of instrument validity.

Step 7: Develop Guidelines for Use

Once it has been determined that there is adequate match between the items and the conceptual framework, guidelines for the use of the scale should be provided. First, a score conversion table should be provided so researchers and practitioners can compare students' raw scores from the survey to their corresponding logit scores in the table. From there, the researcher or practitioner can go to the variable map and plot where a student is compared to the items and construct definition. Finally, the variable map should be formatted and explicated in such a way that practitioners and researchers can easily discern the meaning of the construct and how their participants align to the conceptual definition of the scale.

Data Collection

Data was collected from two sources: de-identified data sets from the *Supporting Success in Algebra* study and cognitive interviews with high school students.

Secondary Data from the *Supporting Success in Algebra* Study

Self-reported survey data was made available from the *Supporting Success in Algebra* study. In the spring of 2018, surveys were administered in two Boston-area high schools to pilot the Comparative Mathematics Anxiety Scale items, as well as items for the other mathematics attitudes of interest in the study. In the fall of 2018, surveys were administered to districts nationwide. In this administration, three forms were created—each form included one of the three Comparative Mathematics Anxiety subscales and all of the other mathematics attitude scales of interest. Data was collected using paper-and-pencil surveys and Scantron forms in spring and fall 2018. The final administration occurred in fall 2020 at one Boston-area high school; because of COVID-19, the survey was administered electronically using the Qualtrics survey platform. This survey included the final version of the Comparative Mathematics Anxiety Scale and the validation scales. Each of these scales is described in more detail below.

Comparative Mathematics Anxiety Scale items. The items in the Comparative Mathematics Anxiety Scale are third-person statements that measure each of the three dimensions of mathematics anxiety: emotional-attitudinal, mental-cognitive, and physical-somatic mathematics anxiety. Students are asked to compare themselves to the student described in each item. Specifically, the instructions for the emotional-attitudinal mathematics anxiety scale are: "Each of the statements below provides a scenario describing how a student feels when working on math problems. Please take a moment to reflect on how you typically feel when working on a set of math problems. Then, compare how you feel to how each student feels." Response options are: much more anxious than [the student described], more anxious than, as anxious as, less anxious than, and much less anxious than. The instructions for the mental-cognitive and physical-somatic scales were adapted to their context, from "student feels" to "student thinks" and "student feels physically", respectively.

The modified Abbreviated Mathematics Anxiety Scale (mAMAS). The modified AMAS was developed by Carey, Hill, Devine, and Szűcs in 2017 to meet the need for a brief measure of mathematics anxiety that is suitable for children and adolescents. As mentioned previously, the AMAS is an abbreviated version of Richardson and Suinn's (1972) Mathematics Anxiety Rating Scale (MARS) created by Hopko, Mahadevan, Bare, and Hunt (2003) both of which are appropriate for adult populations. The mAMAS shows evidence of good reliability (ordinal alpha of 0.89 for the entire scale and 0.83 for both subscales), as well as evidence of construct, convergent, and divergent validity (Carey et al., 2017).

Other measures of student attitudes toward mathematics. The *Student Opinions Toward Mathematics* instrument was developed at Boston College in coordination with Education Development Center as part of the *Supporting Success in Algebra* study [NSF 1621011]. The instrument was piloted in spring 2018 and then administered in fall 2018 and spring 2019, with minor revisions made for each administration. Based on the results, a final version was administered as part of the fall 2020 survey that included the following mathematics attitude constructs, in addition to the mAMAS anxiety scale: confidence, growth and fixed mindset, motivation, and perseverance when doing mathematics.

Demographics. Different demographic data was available from each administration. Gender was collected on the spring 2018 pilot survey. However, no demographic data was collected on the fall 2018 survey, as this data was collected through secondary data requests to the study districts and not available for this dissertation. The final survey administered in fall 2020 included grade-level, gender, and perception of English-language proficiency.

Cognitive Interviews

Cognitive interviewing is an "approach to evaluating sources of response error in survey questionnaires" (Willis, 1999, pg. 1). The method is used in survey development and revision in an attempt to understand the cognitive processes that individuals go through when responding to the survey questions. In this way, researchers aim to develop an instrument that will best capture individuals' true value of a trait. There are two main cognitive interviewing methods: think-aloud and verbal probing. With both methods, the interviewer reads each question to the interviewee and then records or otherwise notes how the subject thought about the questions and arrived at their answer. Where they vary is in how the interviewer interacts with the interviewee after this exchange. In the think-aloud method, the interviewer will say little else except something like "tell me what you are thinking" when the interviewee pauses, while under the verbal probing method, the interviewer will follow-up asking for other, specific information relevant to the question or to the answer given (Willis, 1999).

Two advantages of the verbal probing technique are: 1) ease of training the interviewee, especially in contrast to the think-aloud method which places much of the burden on the respondent, and 2) control of the interview (Willis, 1999). This second advantage is probably the biggest, as "the interviewer can focus on particular areas that appear to be relevant as potential sources of response error" (Willis, 1999, p. 6). This is particularly useful when making revisions to an instrument, as is the current situation, because the interviewer can stay concentrated on particular aspects of the instrument that are the focus of instrument revision.

Cognitive interviews were conducted in November 2019 using an adapted version of the verbal probing technique with a convenience sample of five high school freshman. These were conducted to assess high school students' ability to respond to the items with unique third-person response options, particularly in response to the second, fall 2018 administration of the Comparative Mathematics Anxiety Scale (findings to be discussed in the next chapter). The verbal probing technique was adapted such that students were first administered the fall 2018 version of the survey, rather than being read the survey questions aloud and asked to answer in real-time. The next day, the student was interviewed following a protocol that the interviewer used to probe student thinking about the survey as a whole, as well as to specific items based on their responses to the survey. These cognitive interviews served as an opportunity to gauge high school students' ability to understand and respond to these uniquely-formatted items, as well as to the potential need for any item scaffolding.

In addition to the more traditional cognitive interviews described above, another set of cognitive-type interviews were conducted with a different convenience sample of students at the high school. In this set of interviews, students were asked to participate in what could be considered a card-sorting, or in this case item-sorting, activity, where they were provided the set of items from one dimension of the construct on individual slips of paper and asked to sort them along a continuum from low to high mathematics anxiety. Questions were asked during and after the activity based on how they ordered the items, if they grouped any items, and if they seemed to struggle with the task. After the student was done sorting the items, a picture of the result of the activity was taken to capture how the student thought about the construct and ordered the items along a continuum from low to high mathematics anxiety. Time permitting, students were asked to sort the items for more than one dimension of mathematics anxiety. In all, ten students from grades 10-12 participated in the item-sorting activity cognitive interview. Data collected from these interviews were particularly useful for informing the final round of scale and item revisions, as they provided information about how students interpreted entire items, specific words in an item, the items' relationship to each other, and the items' relationship to the constructed of interest.

Data Analysis

Overarching Research Question

The overarching research question for this study is: *To what extent can the construct of mathematics anxiety be measured reliably and meaningfully by developing a Rasch-based scale?* This research question was answered using Rasch rating scale analysis and through qualitative analysis of the cognitive interviews. Rasch measurement theory allows for a family of probabilistic models, where the choice of model depends on how the data are collected and the intended use of the data (Wright & Masters, 1982). The Rasch rating scale model was used to analyze the scale data collected from the pilot phase and full-scale administration. For this instrument, the Rasch rating scale model is the appropriate model as it assumes all items in the scale have the same response options and individuals are expected to use the response options in the same way. The Rasch rating scale model also assumes the response categories are ordered choices (Andrich, 1978; Rasch, 1960/1980; Wright & Masters, 1982) and is represented by the following statistical model:

$$\pi_{nix} = \frac{e^{\sum_{j=0}^{x_{ni}}[\beta_n - (\delta_i + \tau_j)]}}{\sum_{x=0}^{m} e^{\sum_{j=0}^{k}[\beta_n - (\delta_i + \tau_j)]}}$$

where π_{nix} is the probability of person *n* responding in category *x* to item *i*; β is the ability estimate of person *n*; δ is the location/'difficulty' of item *i* on the variable; τ is the 'threshold' parameter (or the location of the k^{th} step in each item relative to that item's scale value); *m* is the highest category; and *k* is the highest threshold. The Rasch rating scale model is appropriate over the Rasch partial credit model because the rating scale model assumes that the scoring categories have the same relationship to one another across all items; this is in contrast to the partial credit model where the scoring categories are *not* expected to have the same relationship across all items (Andrich, 1978; Wright & Masters, 1982). For scales where all items have the same response options and individuals are expected to use the response options in the same way, the Rasch rating scale model is most appropriate.

Under the Rasch rating scale model, person ability and item difficulty are in logits, which is a unit of measurement that allows us to report relative differences in person abilities and item difficulties by putting them on an equal-interval, standardized unit of measurement (Ludlow & Haley, 1995). Person logits are the natural log odds of a person succeeding on an item with a 0 logit difficulty (average difficulty). Therefore, the higher the person logit, the higher is their ability. Or, for this study, the higher the person logit, the higher the person's mathematics anxiety. An individual with a person ability estimate of 0 has average mathematics anxiety. Similarly, item logits are the natural log odds of a person with 0 logit ability (average ability) failing an item. Therefore, the higher the item logit, the harder the item is. Or, for this study, the higher the item logit, the harder it is to say that one is *more anxious* or *much more anxious* than the person described in the item. Individuals with high anxiety estimates are more likely to respond *about the same*, *more anxious*, or *much more anxious* on an item, while individuals with low anxiety estimates are more likely to respond *less anxious* or *much less anxious* on an item. For items that are difficult to endorse, individuals are more likely to select lower response categories such as *less anxious* and *much less anxious*, and for an easy item, individuals are more likely to select higher response categories such as *more anxious* and *much more anxious*. The WINSTEPS software package was used to perform the analyses (Linacre, 2017, v4.0.1).

The results of these analyses include the following: “variable maps,” or graphical representations of the locations of persons and items based on the data; category characteristic curves (CCCs); item and person fit statistics; and item and person separation indices and reliability estimates. A residual analysis was also conducted to assess whether or not there are any non-random components in the standardized residuals (Ludlow, 1986).

First, variable maps are analyzed to look for evidence of construct validity. Specifically, variable maps provide the following, a visual assessment of: 1) the ordering of the items; 2) the spread of the items; 3) the spread of the persons; and 4) what it means (based on the ordering of the items) to move from low to high anxiety. To determine whether the response options function as intended, the CCCs are analyzed next (Ludlow et al., 2014, 2019; Wright & Masters, 1982; Wolfe & Smith, 2007). Specifically, we want to see that: 1) all five response options were used by respondents, shown visually by the dominance of each response option, or category, in one area of the participants’ anxiety level distribution; and 2) the response options are ordered, such that the probability of moving from one category increases in the expected order from much less anxious occurring first to much more anxious coming last. CCCs are presented using Rasch-Andrich category estimates (Andrich, 1978), which represent the logit value at which there is a 50% probability of moving from one category to the next higher one.

Next, item and person fit statistics are presented and analyzed (Wolfe & Smith, 2007). Rasch fit analyses provide both weighted and unweighted fit statistics with corresponding approximate t-statistics (Wright & Masters, 1982). An unweighted fit statistic is the mean square of the standardized residuals between observed and expected responses, while a weighted fit statistic is the mean square of the standardized residuals between observed and expected responses weighted (or multiplied by) the variance of the item (Wright & Masters, 1982).

Unweighted fit statistics are used to identify outlier responses, while weighted fit statistics are used to assess the consistency of responses (Wright & Masters, 1982). For both statistics, values greater than 1.4 are often flagged as indicating at least one highly unexpected response has occurred (Wolfe & Smith, 2007). During initial instrument development, a more liberal threshold of 1.2 or 1.3 is recommended to avoid missing potential problems in the items (Ludlow et al., 2014). Once items and persons are flagged, additional investigation may be required to understand what is contributing to the item or person misfit. Potential contributions may include: the difficulty of the item, the order of the item on the instrument (also known as entry order), and confusing item or instruction wording.

Item and person separation indices and reliability estimates are then presented. All of these provide an estimate of the ratio of true measure variance to observed score variance (Wright & Masters, 1982). The person separation index and person reliability estimate indicate the extent to which the instrument is sensitive enough to differentiate between low-anxiety and high-anxiety respondents, while the item separation index and item reliability estimate indicate the extent to which the items reliably separate and span to define the entire construct continuum (Linacre, 2016). Ideally you have a person separation index greater than 2, a person reliability estimate greater than 0.8, an item separation index greater than 3, and an item reliability estimate greater than 0.9 (Linacre, 2016).

Finally, a residual analysis was conducted on the final round of data (Ludlow, 1983). First, standardized residuals were exported from Winsteps and imported into SPSS (IBM, 2013, v. 22.0). SPSS was then used to perform a principal components analysis on the residuals (Chou & Wang, 2010; Linacre, 1998a, 1998b). The following statistics and figures were reported: the determinant, the Kaiser-Meyer-Olkin (KMO), Bartlett's Test of Sphericity, factor eigenvalues greater than one, and the scree plot. The determinant indicates the amount of variability in the correlation or covariance matrix. A non-zero determinant indicates there is enough variability in the residuals to conduct a factor analysis; therefore, we want to see a determinant of zero, indicating there is no variability in the residuals. The KMO is an estimate of sampling adequacy that ranges from 0 to 1, where 0 indicates the items do not share a common factor and 1 indicates the items do share a common factor. Typically, a KMO value larger than 0.7 is preferred; however, for the residual analysis, smaller KMO values are preferred. Bartlett's Test of Sphericity and corresponding significant value indicate the extent of variation in the items residual correlation matrix. A significant Bartlett's test is typically preferred, but a non-significant Bartlett's test is desired in a residual analysis. The scree plot provides a visual representation of the eigenvalues extracted for each factor. Typically, eigenvalues greater than 1 indicate a meaningful factor; however, it is also important to consider where there is an appreciable drop (or difference) in the factor eigenvalues. The scree plot provides a visualization of where this drop occurs; an elbow shape indicates where the drop occurs and the eigenvalues plateau, indicating these components may not be meaningful. In a residual analysis, the scree plot should be flat (no elbow shape) with zero eigenvalues greater than 1.

Cognitive interviews were analyzed qualitatively. While a standardized process of transcribing and analyzing qualitative data is preferred, it is often not feasible or necessary for instrument development and revision. Instead, it may be sufficient simply to rely on the interviewers' written notes or a summary of such notes (Willis, 1999). Qualitative analysis of the cognitive interviews focused on dominant trends that came up across multiple interviews as well as "discoveries" of problems that may be expected to have a severe impact on data quality, even if the problem only came up in a single interview (Willis, 1999). Specifically, trends of interest may include: incomprehension of terms (words that individuals consistently had trouble with or misinterpreted) and lack of confidence in response selection (individuals consistently had trouble or hesitated in selecting a response option to a specific item, a set of items, or all items). In addition to trends across interviews, unique problems may also be of interest. It is hard to predict unique responses that may come up during the interviews, but focused attention was given in the analysis to surprise answers that may severely impact data quality.

Research Question 1

In addition to answering the overarching research question, the study sub-questions were also answered. Sub-question 1 is: *How can the tripartite conceptual framework of mathematics anxiety Dr. Caroline Vuilleumier and I developed be employed and adapted in the development of a Rasch-based scale*? This question was answered over the course of instrument development by capturing the process and procedures that were used to develop the Rasch-based scale. Data analyses involved synthesis and narrative description of how the scale was developed with the guidance of the new mathematics anxiety conceptual framework but also how the framework may or may not have been adapted to appropriately capture the tripartite conceptual nature of mathematics anxiety as found in the data.

Research Question 2

The second sub-question is: *Is the scale invariant across grade, gender, and other student demographics?* Measurement invariance, or measurement equivalence, is a statistical property of measurement that indicates the same construct is being measured across some specified groups (e.g., male and female students). One data analysis method that is appropriate for assessing the measurement invariance of a scale is differential item functioning (DIF) analysis (Holland & Wainer, 1993; Osterlind & Everson, 2009; Walker, 2011). An item is labeled as having DIF when people (or in this case students) with equal ability (or in this case, mathematics anxiety), but from different groups (e.g., male versus female), have an unequal probability of item success (in this case, unequal probability of selecting a particular response option). In contrast, an item would be labeled as not having DIF in this dissertation when students with equal mathematics anxiety have an equal probability of selecting a particular response option, regardless of their group membership.

DIF analyses for grade, gender, and English language proficiency (ELP) were run using Winsteps software (Linacre, 2017, v4.0.1). Winsteps has two tests of significance for DIF analyses: the Rasch-Welch t-test and the Mantel methods, and both were used in this study (Linacre, n.d.). The Rasch-Welch t-test logistic regression, or logit-difference, method that "estimates the difference between the Rasch item difficulties for the two groups, holding everything else constant" (Linacre, n.d., DIF/DPF Statistics section, para. 1). The Mantel methods, on the other hand, are log-odds estimators from crosstabs of observations for the two or more groups and includes two different tests. The Mantel-Haenszel test is used for dichotomous variables (e.g., gender and ELP), while the Mantel test is used for polytomous variables (e.g., grade level).

Research Question 3

The final research sub-question is: *Does the scale have an acceptable degree of construct and criterion validity?* The construct validity of the scale was partially examined while answering the overarching research question. Specifically, assessing the fit of the variable map output from Winsteps to the *a priori* conceptualization of mathematics anxiety after each round of data collection provides partial evidence for the construct validity of the measure. Additional evidence for the construct validity of the scale comes from the explication of the construct and description for use after the final round. That is, after the last round of data is collected and it has been determined that there is an acceptable degree of match between the conceptual framework and the scale items, the table aligning the logit scores to the raw scores was provided, and the variable map was formatted and explicated so that practitioners and researchers can easily discern the meaning of the construct and how their potential survey takers would align to the conceptual definition of the scale. The degree of criterion validity was determined by looking at the correlation of students' scores on the Rasch-based Comparative Mathematics Anxiety Scale and their scores on the other measures collected in the survey. Logit estimates of students' scores were used for the Comparative Mathematics Anxiety Scale, while raw sum scores were calculated for all other scales after reverse coding any necessary items. IBM's SPSS Statistics software was used to conduct these analyses (IBM, 2013, v. 22.0). Specifically, a positive correlation between student scores on the Rasch-based Comparative Mathematics Anxiety Scale and their scores on the mAMAS (Carey et al., 2017) and the fixed mindset scale provides convergent validity evidence, while negative correlations between student scores on the Rasch-based mathematics anxiety scale and their scores on the following constructs: confidence, growth mindset, motivation, and perseverance when doing mathematics provide divergent validity evidence, all of which provide evidence for the criterion validity of the scale.

Quality Assurance

To ensure the quality of the data and the analyses conducted, procedures consistent with the standards laid out here were followed. Collection, recording, and analyses of data were documented clearly and thoroughly so that they could be reviewed by others and by myself at a later date without confusion. Data collectors were trained in the protocols to be administered for quality assurance. Survey data collected in the *Supporting Success in Algebra* study were reviewed at each stage for discrepancies and anomalies; any discrepancies and anomalies found were corrected. For the collection and analysis of interview data, procedures were followed as closely as possible and notes were taken when adjustments to the protocols were made. Notes and analyses from interviews were reviewed at several time points by myself and at least one other individual to ensure quality of data and appropriateness of the inferences made. At all stages, care was taken to record the process with as much detail and accuracy as possible.

Ethical Considerations

As this data was collected on what some consider a sensitive topic (mathematics anxiety) and because the population of interest is students (some of who were minors), it was important to ensure that the data was collected with the utmost care and sensitivity. The research was approved by an institutional review board (IRB) and all IRB guidelines were followed. All participants were informed of their rights as research participants, and it was made clear that they could choose not to participate at any time (including retracting their participation in the middle or at the end). All primary data was collected only after obtaining student assent and parent consent. Survey responses were obtained as secondary data from the *Supporting Success in Algebra* study, and therefore, already de-identified. All data was maintained on password-protected, external hard drives so as to maintain the strictest standards of data protection.

Ethical considerations were particularly important for the cognitive interviews. While the interviewees could not be anonymous to the interviewer, data collected during the interview process was quickly de-identified and then securely stored. During the interviews, as part of the job of collecting the data, the interviewer monitored the comfort-level of each participant. As determined by the interviewer, more sensitive questions were skipped or an interview was ended if they felt it was necessary. If ethical concerns had come up during the course of the study, they would have been discussed and resolved with the committee immediately upon discovery. No such concerns occurred during the course of this dissertation research.

CHAPTER 4 – RESULTS

Overarching Research Question: Developing a Rasch-based Scale of Mathematics Anxiety

Again, the overarching research question for this study is: *To what extent can the construct of mathematics anxiety be measured reliably and meaningfully by developing a Rasch-based scale?* This research question will be answered through the iterative scale development process that began in the literature review section. Rasch rating scale analysis is being used as the basis for the scale development, with qualitative analysis of the cognitive interviews providing more in-depth data about the survey and the construct. We begin by briefly revisiting the results from the literature review and the implications of those results for next steps.

Revisions: Round 1

Step one of the iterative scale development process is to gather and synthesize available information. As described in the literature review chapter, the mathematics testing anxiety scale developed and administered in Vuilleumier and Klein (2015) did not fully match the theory, as items for all three dimensions did not span all five levels as hypothesized. Two potential reasons for this were proposed: 1) flaws in item writing that resulted in items not capturing all levels, and 2) flaws in the conceptualization of mathematics anxiety such that the three dimensions should not be expected or theorized to span all five levels. To gain clarity on this issue, additional items needed to be developed to try to flesh out the full range of the scale for all three dimensions. If successful, this would suggest the first reason is accurate (the item writing was flawed); if, however, the items still do not span all five levels, this would support the second conclusion (the theory is flawed). In addition, even within dimensions, items were often disordered from the theorized ordering. This suggests the need to revisit the instrument to improve item wording. Finally, the original scale was developed to measure mathematics testing anxiety not general mathematics anxiety and administered to undergraduate students instead of high school students. While I did not anticipate that either of these differences would impact the behavior of the items, it was worth another administration to check this hypothesis. For these two reasons, a quick revision and administration of the scale was in order before making substantial changes to the items or to the theory of mathematics anxiety.

While the items did not fully match our conceptualization of mathematics (testing) anxiety, it was decided not to change the theoretical conceptualization of the construct without further evidence (Step 2). Therefore, in collaboration with content knowledge experts and after a quick review of the literature, the items were revised (Step 3, see *Table 4* on the following page). Specifically, the following changes were made: one item from the mental-cognitive dimension (Sarah has a mental fog) was removed because of inconsistency across data collection samples; most items were slightly revised for clarity or to remove reference to testing/exams; several items in each dimension were re-ordered from their original conceptualization based on evidence from the two rounds of data collection; additional items were written in each dimension to fill in gaps on the scale or extend the range of the scale; and the names of the students in the items were changed to be more universal. Once the revisions were finalized, data was collected from 8th grade and high school students at a district in the greater-Boston area in spring 2018 (Step 4). The results are presented below (Step 5) and allow for an examination of the match between the items and the conceptual framework for mathematics anxiety (Step 6).

Table 4

Comparative Mathematics Anxiety Items, Revisions Round 1 (R1)

Level of Anxiety	Dimension of Mathematics Anxiety		
	Emotional/Attitudinal	*Mental/Cognitive*	*Physical/Somatic*
Overwrought	1_5. Ruby is terrified. 1_9. Isaac feels intense fear. 1_13. Larissa feels emotionally worked up.	2_6. Jaime thinks about how anxious he is instead of the problem in front of him. 2_9. Emilie can think about nothing else except her anxiety in math.	3_3. Laila gets nauseous to the point of feeling like she is going to be sick. 3_11. Martin is on the verge of tears.
Distressed	1_1. Michael is completely on edge. 1_2. Sebastian feels extremely overwhelmed. 1_10. Andrea is distressed.	2_1. Victor's mind completely freezes. 2_4. Aurora is easily distracted.	3_1. Ivan feels physically uncomfortable. 3_6. Jasper gets a stomachache. 3_10. Erica gets a headache.
Antsy	1_6. Alicia feels uneasy. 1_12. Gabriel has a sense of worry.	2_2. Ana can't recall what she learned. 2_10. Joseph's mind keeps blanking out.	3_2. Gabrielle's heart is racing. 3_5. Angela bites her nails. 3_12. Liam cannot stop squirming in his seat.
Apprehensive	1_3. Viktoria is a bit apprehensive. 1_7. David feels a little intimidated.	2_5. Petra feels confused. 2_8. Leon's thoughts are scattered. 2_11. Lucas feels like he should know how to solve the problem but he can't put his thoughts in order.	3_7. Nora gets a little tense. 3_9. Noah's palms start to sweat.
Tranquil	1_4. Daniel feels in his element. 1_8. Sarah is excited. 1_11. Chloe feels unfazed.	2_3. Carlos is able to think clearly. 2_7. Lena is focused on the problems in front of her.	3_4. Jay is completely calm. 3_8. Romita feels completely at ease.

Results: Round 1

Emotional-attitudinal scale. The variable map for the emotional-attitudinal scale from the first round of data collection can be found in *Figure 4* on the following page. Overall, the persons appear to be fairly well spread, but the items are not spreading to the lowest and highest ends of the continuum. We can confirm this by looking at the person and item separation indices and reliability estimates. Again, the person separation index and reliability estimate indicate the extent to which the instrument is sensitive enough to differentiate low- and high-anxiety respondents, while the item separation index and reliability estimate indicate the extent to which the items reliably separate and span to define the entire construct continuum. Ideally, we see a person separation index greater than 2, a person reliability estimate greater than 0.8, an item separation index greater than 3, and an item reliability estimate greater than 0.9 (Linacre, 2016). Here, the person separation index is 2.10 and the person reliability is 0.81 therefore meeting the criteria, while the item separation index is 2.48 and the item reliability estimate is 0.86, both of which are too low to meet the criteria. Finally, when comparing the ordering of the items in the variable map to the expected ordering in *Table 4*, we notice some disordering of the items. Specifically, Larissa (A1_13), Sebastian (A1_2), & Andrea (A1_10) are falling lower (easier) than theorized, while Alicia (A1_6) appears slightly higher (harder) than expected.

Figure 4

Comparative Math Anxiety (Emotional-Attitudinal) Variable Map, R1

```
MEASURE PERSON - MAP - ITEM
         <more>|<rare>
    2          +
               |
               |
               |
               |
               |
            X  |
               |
              T|
          XXX  |
    1          +
               |
            X  |T
               |  A1_9_Isaac
         XXXX  |  A1_5_Ruby
         XXXX  |
            X  |S
          XXX S|  A1_13_Larissa   A1_1_Michael
            X  |
         XXXX  |  A1_6_Alicia
    0     XXX  +M A1_10_Andrea    A1_12_Gabriel
           XX  |  A1_2_Sebastian  A1_7_David
               |  A1_3_Viktoria
          XXX  |
          XXX  |S A1_8_Sarah
            X M|  A1_11_Chloe
        XXXXX  |  A1_4_Daniel
        XXXXX  |
        XXXXX  |T
       XXXXXX  |
   -1  XXXXXX  +
          XXX  |
           XX  |
           XX S|
         XXXX  |
               |
               |
            X  |
               |
               |
   -2      XX  +
              T|
               |
               |
               |
               |
               |
            X  |
            X  |
               |
   -3          +
         <less>|<freq>
```

Next, we look at the category characteristic curves (CCCs) shown in *Figure 5* to determine whether the response options function as intended. We can see that (1) all five response options were used by respondents, shown visually by the dominance of each response option in one area of the participants' anxiety level distribution; and (2) the response options are ordered, such that the probability of moving from one category increases in the expected order from much less anxious to much more anxious. While both of these are true, the curves for the third and fourth response options are relatively flat, and it would be better to see those dominate a bit more in their portion of the anxiety distribution.

Figure 5

Comparative Math Anxiety (Emotional-Attitudinal) CCCs, R1

```
        CATEGORY PROBABILITIES: MODES - Andrich thresholds at intersections
P      -+---------+---------+---------+---------+---------+---------+-
R  1.0 +                                                             +
O      |                                                             |
B      |                                                             |
A      |1                                                         555|
B   .8 + 111                                                    55   +
I      |    111                                              555     |
L      |       11                                          55        |
I      |         11                                      55          |
T   .6 +           1                                    5            +
Y      |            11                                55             |
    .5 +              1                              5               +
O      |               11                          55                |
F   .4 +               22**222222                 5                  +
       |           2222    1     2**3333333*4444**44444              |
R      |         22         11 333  22   44 3335       4444          |
E      |      222            3*       2*4    5533          444       |
S   .2 +  2222            333  11   444 22  5    33           444    +
P      |22              33       1*4     5**       333           4444|
O      |            3333       444 111 55   222       333            |
N      |     3333333     444444   5555*1111    222222    3333333     |
S   .0 +*****************555555555         1111111111****************+
E      -+---------+---------+---------+---------+---------+---------+-
       -3        -2        -1         0         1         2         3
        PERSON [MINUS] ITEM MEASURE
```

After looking at the variable map and the CCCs, we next want to review the item and person misfit tables (*Tables 5* and *6*, respectively). As a reminder, both tables provide weighted (INFIT) and unweighted (OUTFIT) fit statistics, where INFIT statistics assess the consistency of responses and OUTFIT statistics identify outlier responses. For both statistics, MNSQ values greater than 1.4 should be flagged as indicating at least one highly unexpected response has occurred. During initial instrument development, a more liberal threshold of 1.2 or 1.3 is recommended to avoid missing potential problems in the items.

Table 5

Comparative Math Anxiety (Emotional-Attitudinal) Item Fit Statistics, R1

Item	Measure	INFIT		OUTFIT	
		MNSQ	*ZSTD*	*MNSQ*	*ZSTD*
A1_8	-0.39	2.48	7.3	2.54	7.3
A1_9	0.65	1.88	4.3	1.52	2.6
A1_4	-0.60	1.58	3.4	1.68	3.8
A1_11	-0.55	1.31	2.0	1.36	2.2
A1_5	0.58	1.24	1.5	1.08	0.5
A1_13	0.34	0.93	-0.4	0.87	-0.8
A1_2	-0.14	0.79	-1.4	0.77	-1.5
A1_10	0.03	0.76	-1.7	0.75	-1.7
A1_7	-0.09	0.56	-3.4	0.59	-3.1
A1_1	0.30	0.58	-3.1	0.53	-3.5
A1_3	-0.21	0.51	-4.1	0.55	-3.5
A1_12	-0.02	0.46	-4.5	0.49	-4.0
A1_6	0.08	0.41	-5.0	0.43	-4.6

From *Table 5*, we see that five items (A1_8, A1_9, A1_4, A1_11, and A1_5) have an INFIT MNSQ greater than 1.2, and four items (the same items, excluding item A1_5) have an OUTFIT MNSQ greater than 1.2. Notice that they are the three easiest (A1_8, A1_4, and A1_11) and two hardest (A1_9 and A1_5) items on the emotional-attitudinal mathematics anxiety scale. *Table 6*, on the following page, provides the statistics for students with significant misfit. Fourteen students had INFIT and OUTFIT MNSQ values greater than 1.4, and an additional five students had INFIT and/or OUTFIT MNSQ values greater than 1.2 but less than 1.4.

Table 6

Comparative Math Anxiety (Emotional-Attitudinal) Person Fit Statistics, R1

Person	Measure	INFIT		OUTFIT	
		MNSQ	*ZSTD*	*MNSQ*	*ZSTD*
74	0.41	3.31	4.4	3.51	4.7
68	0.03	3.35	4.5	3.45	4.6
73	0.65	3.02	4.0	3.20	4.2
1	0.49	2.94	3.9	3.15	4.2
78	-1.07	2.80	3.3	2.42	2.8
76	0.26	2.61	3.4	2.65	3.5
72	0.49	2.36	3.0	2.51	3.3
20	0.82	2.12	2.6	2.49	3.1
63	1.09	2.27	2.7	2.39	2.8
47	0.11	2.26	2.9	2.35	3.0
71	0.57	2.09	2.6	2.06	2.5
69	-0.60	2.05	2.4	1.90	2.1
55	0.34	1.52	1.4	1.56	1.5
65	-1.41	1.46	1.1	1.41	1.0
12	0.34	1.34	1.0	1.38	1.1
6	-2.71	1.33	0.7	1.04	0.3
44	0.11	1.28	0.9	1.31	0.9
31	-1.17	1.26	0.7	1.25	0.7

To better understand why this misfit is occurring, we can look at the table of poorly fitting persons (Table 7.1) in Winsteps, which shows the observed person (or student) response, the expected response, and the standardized residual (z-residual) for the students with the greatest misfit across all of the items (INFIT and/or OUTFIT MNSQ greater than or equal to 1.2). The observed responses, expected responses, and z-residuals for the five students with the highest misfit are provided in *Table 7* on the following page.

Table 7

Comparative Math Anxiety (Emotional-Attitudinal) Poorly Fitting Persons, R1

Person	Response Statistic	Item												
		1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*
74	Obs	5	4	4	1	4	3	3	1	5	5	1	3	5
	Exp	3.1	3.5	3.6	4.0	2.8	3.3	3.5	3.8	2.7	3.4	3.9	3.4	3.0
	Z				-3				-2	2		-3		
68	Obs	3	5	1	1	5	3	3	1	5	4	1	3	4
	Exp	2.7	3.1	3.2	3.6	2.4	2.9	3.1	3.4	2.3	3.0	3.6	3.0	2.6
	Z			-2	-2	2			-2	2		-2		
73	Obs	5	5	3	2	5	3	3	1	5	5	2	3	5
	Exp	3.3	3.8	3.8	4.2	3.0	3.6	3.7	4.0	3.0	3.6	4.1	3.7	3.3
	Z				-2				-3			-2		
1	Obs	4	5	3	1	5	3	4	1	5	5	2	3	4
	Exp	3.2	3.6	3.7	4.0	2.9	3.4	3.6	3.9	2.8	3.4	4.0	3.5	3.1
	Z				-3	2			-3	2		-2		
78	Obs	1	1	2	5	1	1	1	5	1	1	5	1	1
	Exp	1.8	2.1	2.1	2.5	1.6	1.9	2.0	2.3	1.6	1.9	2.4	2.0	1.7
	Z				2				2			2		

Note. Obs = observed response, Exp = expected response, and Z = z-residual.

The student with the highest misfit (74) has significant misfit, according to the z-residual, on items A1_4, A1_8, A1_9, and A1_11. For items A1_4, A1_8, and A1_11, the student was expected to give a response of approximately 4 “more anxious than” but gave a response of 1 “much less anxious than” in all three cases. For the easiest items, the student was giving almost the exact opposite response of what was expected and appeared to be matching the low anxiety item to the low anxiety response, rather than comparing themselves. A similar pattern is seen for many of the other students with misfit (i.e., students 68, 73, 1, 72, 20, and 47). For item A1_9, student 74 was expected to give a response of approximately 3 “as anxious as” but responded 5 “much more anxious than.” Given item A1_9 is the hardest item, it appears this student may have again been matching the item to the response option rather than comparing themselves to the item using the response options. This is a pattern we are likely to continue seeing with the hardest and easiest items. That is, because these items are at the far ends of the continuum, when students gave highly unexpected responses (e.g., if they are matching instead of comparing the items and response options), the items are more likely to have higher z-residuals and therefore higher MISFIT across all of the students.

Mental-cognitive scale. The variable map for the mental-cognitive scale from the first round of data collection can be found in *Figure 6* on the following page. Similar to the emotional-attitudinal scale, the persons appear to be fairly well spread, but the items are not spreading to the lowest and highest ends of the continuum, instead clustering pretty tightly together in the middle of the continuum. We can confirm this by looking at the person and item separation indices and reliability estimates. Here, the person separation index is 1.92 and the person reliability is 0.79 which are slightly too low to meet the recommended criteria; in addition, the item separation index is 1.58 and the item reliability estimate is 0.71, both of which are way too low to meet the criteria. In both cases, the mental-cognitive scale is performing worse than the emotional-attitudinal scale. Finally, when comparing the ordering of the items in the variable map to the expected ordering in *Table 4*, we notice that there is one item that is disordered, Aurora (A2_4), which is falling quite a bit lower than theorized.

Figure 6

Comparative Math Anxiety (Mental-Cognitive) Variable Map, R1

```
MEASURE PERSON - MAP - ITEM
         <more>|<rare>
    2          +
               |
               |
               |
            X  |
               |
              T|
               |
    1     XXX  +
            X  |
               |
          XXX  |
           XX  |T A2_9_Emilie
          XXX S|  A2_6_Jaime
         XXXX  |S A2_1_Victor
        XXXXX  |  A2_10_Joseph
    0    XXXX  +M A2_2_Ana
       XXXXXX  |  A2_11_Lucas   A2_4_Aurora   A2_5_Petra    A2_8_Leon
          XXX  |S A2_7_Lena
        XXXXX  |  A2_3_Carlos
         XXXX M|T
          XXX  |
         XXXX  |
         XXXX  |
   -1       X  +
          XXX  |
          XXX S|
           XX  |
         XXXX  |
          XXX  |
           XX  |
               |
   -2       X  +
              T|
            X  |
               |
               |
               |
               |
               |
   -3          +
            X  |
               |
               |
               |
               |
               |
               |
   -4       X  +
         <less>|<freq>
```

Next, we look at the category characteristic curves (CCCs) shown in *Figure 7* to determine whether the response options function as intended.

Figure 7

Comparative Math Anxiety (Mental-Cognitive) CCCs, R1

```
        CATEGORY PROBABILITIES: MODES - Andrich thresholds at intersections
P      -+---------+---------+---------+---------+---------+---------+-
R  1.0 +                                                             +
O      |                                                             |
B      |                                                             |
A      |                                                           55|
B   .8 +111                                                     555  +
I      |   11                                                 55     |
L      |     11                                             55       |
I      |       11                                         55         |
T   .6 +         11                                      5           +
Y      |           1                                   55            |
    .5 +            11               33333            5              +
O      |              1           333     333       55               |
F   .4 +              2**222222333           33    5                 +
       |          2222   1   33222             33 5                  |
R      |       222        113     222        444***444444            |
E      |    222           3311       22   444  5   3     4444        |
S   .2 + 222           333    11       **4   55     333      444     +
P      |2            33         11  444  22*5          33       44444|
O      |         3333           44**1   555 222          333         |
N      |   333333         444444   55***11     222222       3333333  |
S   .0 +******************555555555       11111111111****************+
E      -+---------+---------+---------+---------+---------+---------+-
       -3        -2        -1         0         1         2         3
        PERSON [MINUS] ITEM MEASURE
```

Again, the curves look pretty good, perhaps even better than they did for the emotional-attitudinal scale. The response options are ordered as they should be. However, we would like to see the second response option and definitely the fourth response option dominating a bit more in their portions of the anxiety distribution.

Finally, we look at the item and person misfit tables (*Tables 8* and *9*, respectively). From *Table 8*, on the following page, we see that three items (A2_3, A2_7, and A2_9) have an INFIT MNSQ greater than 1.2, and two items (the same items, excluding item A2_9) have an OUTFIT

MNSQ greater than 1.2. Notice that they are the two easiest (A2_3 and A2_7) and one hardest (A2_9) item on the mental-cognitive mathematics anxiety scale.

Table 8

Comparative Math Anxiety (Mental-Cognitive) Item Fit Statistics, R1

Item	Measure	INFIT		OUTFIT	
		MNSQ	*ZSTD*	*MNSQ*	*ZSTD*
A2_3	-0.33	2.05	5.4	2.09	5.5
A2_7	-0.26	1.49	2.8	1.50	2.9
A2_9	0.55	1.21	1.3	1.11	0.7
A2_6	0.35	1.12	0.8	1.06	0.4
A2_1	0.24	1.07	0.5	1.06	0.5
A2_10	0.07	0.95	-0.3	0.91	-0.5
A2_4	-0.19	0.71	-2.0	0.73	-1.9
A2_5	-0.15	0.72	-1.9	0.70	-2.1
A2_11	-0.19	0.7	-2.2	0.67	-2.4
A2_2	0.02	0.56	-3.3	0.56	-3.3
A2_8	-0.10	0.56	-3.4	0.56	-3.4

From *Table 9*, on the following page, we see eighteen students had INFIT and OUTFIT MNSQ values greater than 1.4, and an additional two students had INFIT and OUTFIT MNSQ values greater than 1.2 but less than 1.4.

Table 9

Comparative Math Anxiety (Mental-Cognitive) Person Fit Statistics, R1

Person	Measure	INFIT		OUTFIT	
		MNSQ	*ZSTD*	*MNSQ*	*ZSTD*
76	0.28	3.50	4.0	3.53	4.0
73	1.00	3.17	4.2	3.28	4.3
78	-1.79	3.22	3.6	2.84	3.1
72	-0.58	2.97	3.3	3.03	3.4
63	0.59	2.96	3.6	2.98	3.6
20	1.00	2.42	3.1	2.49	3.2
3	-0.47	2.39	2.6	2.37	2.6
65	-0.36	2.22	2.3	2.24	2.3
61	-0.14	2.22	2.3	2.20	2.3
32	-0.14	1.92	1.8	1.89	1.8
40	0.28	1.83	1.8	1.85	1.8
68	0.07	1.79	1.7	1.80	1.7
19	-0.03	1.67	1.5	1.67	1.5
55	-0.03	1.58	1.3	1.60	1.3
56	0.49	1.58	1.4	1.57	1.4
71	-1.47	1.5	1.3	1.57	1.4
12	-0.03	1.55	1.3	1.55	1.3
41	-0.03	1.41	1.0	1.42	1.0
51	-0.47	1.33	0.9	1.33	0.9
69	0.39	1.24	0.7	1.23	0.7

Again, looking at the poorly fitting persons in *Table 10*, on the following page, we continue to see some of the misfit patterns found in the emotional-attitudinal dimension of mathematics anxiety. Student 76 has significant misfit (z-residual greater than or equal to 2) on six items, including the three items that were flagged as having significant misfit in *Table 8*. This student, and others, provided significantly lower than expected responses to the easiest items and significantly higher than expected responses to the harder item(s), suggesting a continued pattern for some students of matching rather than comparing items to the response options. Again, this phenomenon impacts the items at the end of the continuum (the hardest and the easiest items) more so than the items in the middle.

Table 10

Comparative Math Anxiety (Mental-Cognitive) Poorly Fitting Persons, R1

Person	Response Statistic	Item										
		1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*
76	Obs	5	4	1	4	4	5	1	3	5	1	2
	Exp	3.0	3.2	3.5	3.3	3.3	2.9	3.4	3.3	2.7	3.1	3.3
	Z	2		-2			2	-2		2	-2	
73	Obs	5	5	1	3	4	5	1	4	5	5	4
	Exp	3.6	3.8	4.1	4.0	4.0	3.5	4.0	3.9	3.3	3.8	4.0
	Z			-3				-3				
78	Obs	1	1	5	4	1	1	1	1	1	1	1
	Exp	1.5	1.6	1.8	1.7	1.7	1.5	1.8	1.7	1.4	1.6	1.7
	Z			3	2							
72	Obs	5	2	1	1	3	5	1	2	3	2	2
	Exp	2.3	2.4	2.7	2.6	2.6	2.2	2.7	2.5	2.0	2.4	2.6
	Z	3					3					
63	Obs	4	5	1	3	2	5	1	4	5	4	4
	Exp	3.2	3.4	3.8	3.6	3.6	3.1	3.7	3.5	3.0	3.4	3.6
	Z			-2				-2		2		

Note. Obs = observed response, Exp = expected response, and Z = z-residual.

Physical-somatic scale. The variable map for the physical-somatic scale from the first round of data collection can be found in *Figure 8*, on the following page. Notice that the persons appear to be fairly well spread with some clustering, or lack of spread, in the lower half; similarly, the items are spread relatively well on the top half of the continuum (above the mean) but do not spread to the low end of the continuum. Looking at the separation indices and reliability estimates, we see that the person separation index is 1.75, the person reliability is 0.75, the item separation index is 2.40, and the item reliability estimate is 0.85, all of which are too low to meet the suggested criteria. Finally, when comparing the ordering of the items in the variable map to the expected ordering in *Table 4*, we should note that there is some disordering with Jasper (A3_6), Gabrielle (A3_2), and Noah (A3_9) all slightly higher than theorized—each is off by one level.

Figure 8

Comparative Math Anxiety (Physical-Somatic) Variable Map, R1

```
MEASURE PERSON - MAP - ITEM
         <more>|<rare>
    2          +
               |
               |
               |
               |
               |
               |
               |
               |
            X  |
    1          +
               |
            X T|
            X  |T
          XXX  |
           XX  |  A3_3_Laila      A3_6_Jasper
               |S
               |  A3_11_Martin    A3_2_Gabrielle
           XX  |  A3_1_Ivan
           XX  |
    0     XXX S+M A3_10_Erica
          XXX  |  A3_12_Liam      A3_5_Angela     A3_9_Noah
           XX  |
            X  |  A3_7_Nora
        XXXXX  |S
         XXXX  |  A3_8_Romita
        XXXXX  |
           XX M|T A3_4_Jay
          XXX  |
        XXXXX  |
   -1  XXXXXX  +
       XXXXXX  |
         XXXX  |
            X  |
          XXX S|
               |
         XXXX  |
           XX  |
               |
          XXX  |
   -2          +
            X  |
              T|
               |
               |
               |
               |
               |
            X  |
               |
   -3       X  +
         <less>|<freq>
```

Next, we look at the category characteristic curves (CCCs) shown in *Figure 9* to determine whether the response options function as intended. Note that while the response options appear to be ordered, we are not seeing the second, third, and fourth options dominate over a significant portion of the anxiety distribution. More work is needed on these items for the response options to be utilized as intended.

Figure 9

Comparative Math Anxiety (Physical-Somatic) CCCs, R1

```
        CATEGORY PROBABILITIES: MODES - Andrich thresholds at intersections
P      -+---------+---------+---------+---------+---------+---------+-
R  1.0 +                                                             +
O      |                                                             |
B      |1                                                           5|
A      | 1111                                                   5555 |
B   .8 +     11                                              555     +
I      |       111                                         55        |
L      |          11                                     55          |
I      |            1                                  55            |
T   .6 +             11                               5              +
Y      |               11                           55               |
    .5 +                 1                         5                 +
O      |                  1                      55                  |
F   .4 +                   11                   5                    +
       |                 2222*2222333333333 444*4444                 |
R      |             2222     113322     44*3*5     4444             |
E      |          222        3331   22244   5 33        444          |
S   .2 +      2222         33    11 44422 55    33         444       +
P      |  2222          333       4*1   5*2       333         44444  |
O      |22           333       444   ***   222       333           44|
N      |      3333333     44444  5555   1111  222222    3333333      |
S   .0 +******************5555555           11111111*****************+
E      -+---------+---------+---------+---------+---------+---------+-
       -3        -2        -1         0         1         2         3
        PERSON [MINUS] ITEM MEASURE
```

Finally, *Table 11* and *12* show the person and item misfit statistics, respectively. From *Table 11*, on the following page, we see that four items (A3_4, A3_8, A3_3, and A3_11) have an INFIT MNSQ greater than 1.2, and two items (the same items, excluding items A3_3 and A3_11) have an OUTFIT MNSQ greater than 1.2. The items with the highest misfit, where both

INFIT and OUTFIT MNSQ are greater than 1.2, are the two easiest (A3_4 and A3_8), and the two items with just OUTFIT MNSQ greater than 1.2 are the two theorized to be the hardest (A3_3 and A3_11) on the physical-somatic mathematics anxiety scale. We already saw that item A3_6 (Jasper) was unexpectedly harder than item A3_11 (Martin).

Table 11

Comparative Math Anxiety (Physical-Somatic) Item Fit Statistics, R1

Item	Measure	INFIT		OUTFIT	
		MNSQ	*ZSTD*	*MNSQ*	*ZSTD*
A3_4	-0.65	2.23	6.4	2.75	8.1
A3_8	-0.54	2.16	6.0	2.53	7.3
A3_3	0.48	1.36	1.9	1.05	0.3
A3_11	0.33	1.36	2.0	1.12	0.7
A3_5	-0.12	0.74	-1.9	0.71	-2.0
A3_2	0.28	0.72	-1.9	0.68	-2.0
A3_6	0.47	0.72	-1.7	0.59	-2.5
A3_9	-0.06	0.68	-2.3	0.71	-1.9
A3_10	-0.02	0.64	-2.6	0.65	-2.4
A3_12	-0.07	0.65	-2.6	0.61	-2.8
A3_1	0.23	0.56	-3.2	0.57	-2.9
A3_7	-0.32	0.52	-3.9	0.53	-3.6

From *Table 12*, on the following page, we see sixteen students had INFIT and/or OUTFIT MNSQ values greater than 1.4, and an additional four students had INFIT and/or OUTFIT MNSQ values greater than 1.2 but less than 1.4. As explained in the previous two sections, some misfit is not unexpected with these items given the unique response options and the likelihood the misfit from the unique response options impacts the extreme ends of the scale more.

Table 12

Comparative Math Anxiety (Physical-Somatic) Person Fit Statistics, R1

Person	Measure	INFIT		OUTFIT	
		MNSQ	*ZSTD*	*MNSQ*	*ZSTD*
20	1.07	3.43	3.9	4.90	5.2
74	0.78	2.89	3.5	3.75	4.5
73	0.62	2.80	3.5	3.38	4.2
76	0.62	2.82	3.5	3.37	4.2
72	0.08	2.81	3.7	2.91	3.8
63	0.23	2.53	3.2	2.73	3.5
3	-0.14	2.51	3.2	2.58	3.3
47	0.46	2.15	2.6	2.43	3.0
1	0.15	1.94	2.2	2.11	2.5
12	-0.14	2.03	2.4	2.07	2.5
78	-1.42	2.02	1.9	1.51	1.1
64	-0.60	1.98	2.2	1.97	2.2
75	-0.60	1.98	2.2	1.97	2.2
61	-0.29	1.87	2.1	1.88	2.1
44	0.62	1.57	1.5	1.83	1.9
69	-1.29	1.42	1.0	1.25	0.7
58	0.70	1.26	0.8	1.24	0.7
37	-2.13	1.24	0.6	0.85	0.0
57	0.01	1.16	0.6	1.23	0.7
31	-0.69	1.22	0.7	1.20	0.6

Table 13, on the following page, provides the observed responses, expected responses, and z-residuals of each item for the students with the greatest misfit. As with the emotional-attitudinal and mental-cognitive dimensions of mathematics anxiety, we see similar patterns where the greatest item misfit is occurring at the ends of the item continuum with students giving highly unexpected responses to the hardest and easiest items on the scale.

Table 13

Comparative Math Anxiety (Physical-Somatic) Poorly Fitting Persons, R1

Person	Response Statistic	Item											
		1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*
20	Obs	5	5	5	1	5	5	4	1	4	4	5	5
	Exp	3.9	3.9	3.7	4.5	4.2	3.7	4.4	4.5	4.2	4.1	3.8	4.2
	Z				-5				-4				
74	Obs	4	4	5	1	4	5	5	1	3	4	5	5
	Exp	3.6	3.6	3.3	4.4	4	3.4	4.2	4.3	3.9	3.9	3.5	3.9
	Z				-4				-4				
73	Obs	4	4	5	1	4	5	3	1	3	4	5	5
	Exp	3.4	3.4	3.2	4.3	3.8	3.2	4	4.2	3.8	3.7	3.3	3.8
	Z				-3				-3				
76	Obs	2	4	5	1	5	5	4	1	4	4	5	4
	Exp	3.4	3.4	3.2	4.3	3.8	3.2	4	4.2	3.8	3.7	3.3	3.8
	Z				-3				-3				
72	Obs	3	4	5	1	3	3	4	1	2	5	5	1
	Exp	2.8	2.8	2.5	3.8	3.2	2.5	3.5	3.7	3.2	3.1	2.7	3.2
	Z			2	-2				-2			2	-2

Note. Obs = observed response, Exp = expected response, and Z = z-residual.

Revisions: Round 2

At this point, there are two sets of results: the results from the original rounds of data collection with the mathematics testing anxiety scale, as well as the new results from the first round of revisions for the Comparative Mathematics Anxiety Scale. These serve as the available information (Step 1). It is at this point where I revisited the conceptual framework and considered whether or not to make changes to the framework (Step 2). Because the items across all three components of the scale (emotional-attitudinal, mental-cognitive, and physical-somatic) are still relatively unstable, I determined it was not appropriate to make revisions to the conceptual framework at this point. If item locations become more consistent and some items continue to not align with the conceptual framework, then I would consider changes to the conceptual framework.

Based on the data available, revisions were made to items in two of the three dimensions of the Comparative Mathematics Anxiety Scale: the emotional-attitudinal subscale and the mental-cognitive subscale (Step 3, see *Table 14* on the following page). At this time, no changes were made to the items in the physical-somatic subscale, as likely changes to the scale would occur by re-ordering some of the items and therefore impacting the conceptualization of the construct, rather than changing item wording. Therefore, additional evidence of and consistency in item locations would be desirable before making those changes.

In the emotional-attitudinal subscale, item A1_8 (Sarah is excited) was removed as it had the highest INFIT and OUTFIT MNSQ (2.48 and 2.54, respectively). In addition, after reflecting on the wording of the item and the conceptualization of mathematics anxiety, it was determined that the item is somewhat construct irrelevant, more closely measuring enjoyment of mathematics rather than a low level of mathematics anxiety. In the mental-cognitive subscale, item A2_4 (Aurora is easily distracted) was also deleted for two reasons: (1) it was the only item that was disordered from the theorized ordering, and (2) it was also determined to be construct irrelevant, measuring attention rather than anxiety in mathematics. In addition to removing item A2_4, the two hardest items (A2_6 and A2_9) in the mental-cognitive scale were revised in an attempt to make the items harder, that is more difficult for students to select "more anxious" or "much more anxious" than the person described in the item, and spread out the item continuum. Item A2_6 was revised from "Jaime thinks about how anxious *he is* instead of the problem in front of him" to "Jaime thinks about how anxious *math makes him feel* instead of the problem in front of him," and item A2_9 from "Emilie can think about nothing else except her *anxiety in* math" to "Emilie can think about nothing else except her *fear of* math" (emphasis added).

Table 14

Comparative Mathematics Anxiety Items, Revisions Round 2 (R2)

Level of Anxiety	Dimension of Mathematics Anxiety		
	Emotional/Attitudinal	*Mental/Cognitive*	*Physical/Somatic*
Overwrought	1_5. Ruby is terrified. 1_9. Isaac feels intense fear. 1_13. Larissa feels emotionally worked up.	2_6. Jaime thinks about how anxious **math makes him feel** instead of the problem in front of him. 2_9. Emilie can think about nothing else except her **fear of** math.	3_3. Laila gets nauseous to the point of feeling like she is going to be sick. 3_11. Martin is on the verge of tears.
Distressed	1_1. Michael is completely on edge. 1_2. Sebastian feels extremely overwhelmed. 1_10. Andrea is distressed.	2_1. Victor's mind completely freezes. ~~2_4. Aurora is easily distracted.~~	3_1. Ivan feels physically uncomfortable. 3_6. Jasper gets a stomachache. 3_10. Erica gets a headache.
Antsy	1_6. Alicia feels uneasy. 1_12. Gabriel has a sense of worry.	2_2. Ana can't recall what she learned. 2_10. Joseph's mind keeps blanking out.	3_2. Gabrielle's heart is racing. 3_5. Angela bites her nails. 3_12. Liam cannot stop squirming in his seat.
Apprehensive	1_3. Viktoria is a bit apprehensive. 1_7. David feels a little intimidated.	2_5. Petra feels confused. 2_8. Leon's thoughts are scattered. 2_11. Lucas feels like he should know how to solve the problem but he can't put his thoughts in order.	3_7. Nora gets a little tense. 3_9. Noah's palms start to sweat.
Tranquil	1_4. Daniel feels in his element. ~~1_8. Sarah is excited.~~ 1_11. Chloe feels unfazed.	2_3. Carlos is able to think clearly. 2_7. Lena is focused on the problems in front of her.	3_4. Jay is completely calm. 3_8. Romita feels completely at ease.

Once the revisions were finalized, data was collected from high school students enrolled in an algebra support course at a district participating in the *Supporting Success in Algebra* study during fall 2018 (Step 4). As a reminder, data was collected separately for each of the three dimensions of mathematics anxiety to reduce the burden on students, resulting in three different forms of the survey and three different sample sizes. The results of the second round of data collection are presented below (Step 5) and provide some evidence for examining the match between the items and the mathematics anxiety conceptual framework (Step 6).

Results: Round 2

Initial variable maps. The Round 2, initial result variable maps for the emotional-attitudinal, mental-cognitive, and physical-somatic dimensions of mathematics anxiety are presented in *Figures 10*, *11*, and *12* on the following pages, respectively. One will notice almost immediately that all of the variables are reversed; that is, the most difficult items that are expected to be at the top of the mathematics anxiety continuum are at the bottom, and the easiest items that are expected to be at the bottom of the mathematics anxiety continuum are at the top. The quickest way to note this is that all of the theoretically easiest items for all of the dimensions (A1_4 & A1_11; A2_3 & A2_7; and A3_4 & A3_8) are at the top of the items in the variable maps rather than at the bottom. The most obvious cause of this was that the response options had accidentally been reverse coded (a 1 was coded as a 5, a 2 was coded as a 4, etc.); however, this hypothesis was very quickly tested and found to be false.

Figure 10

Comparative Math Anxiety (Emotional-Attitudinal) Variable Map, R2 (initial)

```
MEASURE    PERSON - MAP - ITEM
               <more>|<rare>
    2             .  +
                  .  |
                     |
                  .  |
                     |
                     |
                     |
                  .  |
                  . T|
                  .  |
    1             .  +
                 .#  |
                  #  |
               .###  |
                .## S|
                .##  |T
              #####  |  A1_11_Chloe     A1_4_Daniel
              #####  |S
          .########  |  A1_3_Viktoria
             .#####  |  A1_10_Andrea    A1_7_David
    0  .###########  +M A1_12_Gabriel   A1_6_Alicia
              .#### M|  A1_1_Michael
            .######  |  A1_13_Larissa
                .##  |S A1_2_Sebastian  A1_5_Ruby       A1_9_Isaac
                 .#  |
               ####  |T
                .##  |
                 ## S|
                .##  |
                 .#  |
   -1             .  +
                  .  |
                  .  |
                 .# T|
                  .  |
                     |
                  .  |
                     |
                  .  |
                     |
   -2             .  +
                     |
                     |
                  .  |
                     |
                     |
                     |
                  .  |
                     |
                     |
   -3                +
                     |
                     |
                     |
                  .  |
                     |
                     |
                     |
                     |
                     |
   -4           .##  +
               <less>|<freq>
 EACH "#" IS 7: EACH "." IS 1 TO 6
```

Figure 11

Comparative Math Anxiety (Mental-Cognitive) Variable Map, R2 (Initial)

```
MEASURE    PERSON - MAP - ITEM
               <more>|<rare>
    4             .  +
                     |
                     |
                     |
                     |
                     |
                  .  |
    3                +
                     |
                     |
                  .  |
                     |
                     |
                     |
    2                +
                  .  |
                     |
                  .  |
                  . T|
                  .  |
                 .#  |
    1            .#  +
             .#####  |T A2_3_Carlos   A2_7_Lena
               #### S|
               .###  |
        ###########  |S
       .###########  |
           .#######  |
    0  .########### M+M A2_5_Petra
        .##########  |  A2_1_Victor   A2_2_Ana      A2_8_Leon
               .###  |  A2_10_Joseph  A2_11_Lucas   A2_6_Jaime    A2_9_Emilie
              .####  |S
               .### S|
                 .#  |
               .###  |T
   -1            .#  +
                  .  |
                  . T|
                 .#  |
                  .  |
                  .  |
                  .  |
   -2             .  +
                     |
                  .  |
                  .  |
                     |
                     |
                     |
   -3             .  +
                     |
                     |
                     |
                     |
                     |
                     |
   -4            .#  +
               <less>|<freq>
 EACH "#" IS 9: EACH "." IS 1 TO 8
```

Figure 12

Comparative Math Anxiety (Physical-Somatic) Variable Map, R2 (Initial)

```
MEASURE    PERSON - MAP - ITEM
               <more>|<rare>
    2                +
                     |
                     |
                     |
                     |
                     |
                     |
                     |
                     |
                  . T|
    1                +
                     |
                  #  |
                  #  |
                  #  |
                 ##  |
              ##### S|
               .###  |T A3_4_Jay      A3_8_Romita
             .#####  |S
              #####  |  A3_5_Angela   A3_6_Jasper
    0         #####  +M A3_10_Erica   A3_11_Martin  A3_7_Nora
                ###  |  A3_12_Liam    A3_1_Ivan     A3_3_Laila    A3_9_Noah
            #######  |S
              #####  |T
                  . M|  A3_2_Gabriel
               .###  |
                .##  |
                 .#  |
                 ##  |
                ###  |
   -1            ##  +
                  . S|
                .##  |
                  #  |
                     |
                 .#  |
                 .#  |
                     |
                ###  |
                    T|
   -2                +
                  .  |
                     |
                  .  |
                     |
                  .  |
                     |
                     |
                     |
                     |
   -3                +
                  .  |
                     |
                     |
                     |
                     |
                     |
                     |
                     |
                     |
   -4            ##  +
               <less>|<freq>
EACH "#" IS 2: EACH "." IS 1
```

It was then that another hypothesis emerged—some students may have thought they were supposed to *match* the items to the response options rather than *compare* themselves to the items using the response options. For example, item A1_11 (Chloe feels unfazed) is theoretically one of the easiest items. Using the comparative response options, most students would select “as anxious” if they have no or low emotional-attitudinal mathematics anxiety, “more anxious” if they have some or moderate emotional-attitudinal mathematics anxiety, and “much more anxious” if they have high or severe emotional-attitudinal mathematics anxiety. However, if students did not read the directions or misinterpreted the directions, it is feasible that in reading this item they would match it to “less anxious” or “much less anxious” as this item is describing a student with low mathematics anxiety.

There were several reasons this theory emerged: (1) the format of the response options is unique and would have been new to the students leading to the possibility of using them incorrectly, (2) these items came at the end of a relatively long survey where attention may have started to slip and students may not have read the directions, and (3) due to space limitations on the survey and trying to keep from going onto another page, the response options were reformatted and presented in such a way that may have reduced the clarity and not reinforced the intentions behind the directions. If evidence were found to support this theory and more students went the matching route instead of the comparison route, the variable maps would be reversed rather than as expected. Assuming at least some students followed the directions as intended, we should find two groups of respondents in the data—a reversal group that *matched* items to response options and an expected group that *compared* themselves to the items using the response options. This is a theory which was testable by employing cluster analysis followed by Rasch analysis of the data for any clustered groups that were found.

Cluster analysis and Rasch analysis of the clusters. Cluster analysis is an exploratory analysis that "tries to identify homogenous groups of cases if the grouping is not previously known" (Statistics Solutions, n.d.). That is exactly what is needed to identify the possible groups, or clusters, of students in the Round 2 data. Using SPSS, cluster analysis was run with the following choices: Hierarchical clustering method as the overall analysis method since it does not require the analyst to specify the number of clusters in advance and can handle scale data; Squared Euclidean Distance as the distance measure to calculate distances between observations; and Ward's linkage method as the method to link observations together after the distances have been calculated as it aims to maximize the significance of differences between clusters. Using this method, a dendrogram, which is a graphical representation of how the clusters were merged together and can be used to identify the appropriate number of clusters, was generated for each dimension of mathematics anxiety (see *Figures A1-A3* in Appendix A). Looking at each dendrogram, it is apparent that there are two relatively distinct groups of respondents in each set of data, partially confirming our hypothesis. Once this determination was made, a variable indicating each student's group membership was exported and merged into the original data file so that the data could be separated and Rasch analyses run on both groups. This occurred for all three dimensions and therefore all three data files.

As with any other time, the first step in the Rasch analysis was to examine the variable map. To fully confirm our theory, we anticipated that one variable map for each dimension would look similar to the initial variable map in Round 2 (reversed) while the other variable map would be relatively the opposite of the initial variable map (expected). We found this for each dimension—thus confirming the hypothesis.

The reversed cluster variable maps. *Figures 13-15* on the following pages provide the reversed variable maps for the emotional-attitudinal, mental-cognitive, and physical-somatic dimensions of mathematics anxiety, respectively. While the reversed cluster variable maps are not part of the original data analysis plan, they do provide useful information—what order students think the items go in. That is, because we believe and have evidence to support that these students were matching items to the response options, then logic would indicate that the way they match the items to the response options represents their ideas of how the construct, as represented by the items, is defined. If we look at the reversed variable maps in this way, we can learn something about the ordering of the items as they relate to the conceptual framework.

Figure 13

Comparative Math Anxiety (Emotional-Attitudinal) Variable Map, R2 (Reversed)

```
MEASURE    PERSON - MAP - ITEM
               <more>|<rare>
    3             .  +
                     |
                     |
                     |
                     |
                     |
                     |
                     |
                     |
                     |
    2             .  +
                     |
                  .  |
                     |
                     |
                     |
                     |
                  .  |
                  #  |
                  .  |
    1             . T+
                 .#  |
                 .#  |
              .####  |T
               .### S|  A1_11_Chloe
                ###  |  A1_4_Daniel
            .######  |S
             .#####  |  A1_3_Viktoria
            .######  |  A1_10_Andrea
        ########### M|  A1_12_Gabriel   A1_6_Alicia     A1_7_David
    0    .#########  +M
       ############  |
             .#####  |  A1_13_Larissa   A1_1_Michael
             .##### S|
                 .#  |S
                .##  |  A1_2_Sebastian  A1_5_Ruby       A1_9_Isaac
                  #  |
                 .#  |T
                  . T|
                  .  |
   -1             .  +
                     |
                     |
                  .  |
                     |
                     |
                     |
                     |
                     |
                     |
   -2                +
               <less>|<freq>
 EACH "#" IS 5: EACH "." IS 1 TO 4
```

Figure 14

Comparative Math Anxiety (Mental-Cognitive) Variable Map, R2 (Reversed)

```
MEASURE    PERSON - MAP - ITEM
               <more>|<rare>
    5             .  +
                     |
                     |
                     |
                     |
                     |
                  .  |
    4                +
                     |
                     |
                     |
                  .  |
                     |
                     |
    3                +
                     |
                     |
                  .  |
                     |
                  .  |  A2_3_Carlos
                     |T
    2             .  +
                 .# T|  A2_7_Lena
                     |
                ###  |
                 ##  |
                     |
            .###### S|S
    1          .###  +
           .#######  |
            .######  |
         .#########  |
        .########## M|
          .########  |
          .########  |  A2_5_Petra
    0 #############  +M
             .##### S|
             ######  |
               .###  |  A2_11_Lucas   A2_8_Leon
                ###  |  A2_1_Victor   A2_2_Ana      A2_6_Jaime
                  .  |
                  . T|  A2_10_Joseph
   -1             .  +  A2_9_Emilie
               <less>|<freq>
 EACH "#" IS 4: EACH "." IS 1 TO 3
```

Figure 15

Comparative Math Anxiety (Physical-Somatic) Variable Map, R2 (Reversed)

```
MEASURE  PERSON - MAP - ITEM
             <more>|<rare>
    2              +
                   |
                   |  A3_4_Jay
                   |T
                X  |  A3_8_Romita
                   |
                   |
               XX T|
    1          XX  +
                X  |S
                X  |
              XXX S|
              XXX  |
      XXXXXXXXXXX  |
         XXXXXXXX  |  A3_7_Nora
               XX M|
    0   XXXXXXXXX  +M A3_10_Erica     A3_5_Angela
              XXX  |  A3_1_Ivan       A3_6_Jasper
             XXXX  |  A3_9_Noah
           XXXXXX S|
                X  |  A3_12_Liam
                   |  A3_11_Martin
               XX T|
               XX  |S A3_3_Laila
   -1              +  A3_2_Gabrielle
             <less>|<freq>
```

If we reverse the emotional-attitudinal variable map and compare it to the expected ordering of the items, we find that the items order mostly as expected with the exception of three items (A1_10, A1_13, and A1_2). Similarly, the items in the mental-cognitive variable map order mostly as expected when reversed except for items A2_1, A2_6, and A2_10. Finally, we see slightly more disorder in the physical-somatic variable map once reversed with items A3_9, A3_12, A3_2, and A3_10 not ordering as expected. While perhaps this data should not be weighted as strongly as the other pieces of data, it does provide information worth considering when thinking about possible changes to the items and the conceptual framework. Next, we look at the expected results for each of the three dimensions.

Emotional-attitudinal (expected) results. The expected variable map for the emotional-attitudinal scale from the second round of data collection can be found in *Figure 16*, below.

Figure 16

Comparative Math Anxiety (Emotional-Attitudinal) Variable Map, R2 (Expected)

```
MEASURE    PERSON - MAP - ITEM
               <more>|<rare>
    1              +
                   |T
                   |  A1_5_Ruby       A1_9_Isaac
                   |
                   |S
                   |
                  T|  A1_1_Michael    A1_2_Sebastian
    0              +M A1_10_Andrea    A1_13_Larissa   A1_7_David
                .  |  A1_6_Alicia
                #  |  A1_12_Gabriel   A1_3_Viktoria
                .  |S
              .## S|  A1_4_Daniel
          #######  |  A1_11_Chloe
            #####  |T
   -1      ######  +
              .##  |
                # M|
              ###  |
              ###  |
              ###  |
               ##  |
   -2          ##  +
                  S|
               ##  |
                   |
              .##  |
                   |
               .# T|
   -3              +
                   |
                .  |
                   |
                   |
                   |
                   |
   -4           .  +
                   |
                   |
                   |
                   |
                   |
                   |
   -5     .######  +
               <less>|<freq>
EACH "#" IS 2: EACH "." IS 1
```

Overall, the persons appear somewhat spread; however, most students are falling on the bottom half of the mathematics anxiety continuum and not spreading across the entire continuum. In almost the reverse, the items are somewhat spread out but at the top half and not inclusive of the lower range of the mathematics anxiety continuum. We can confirm this by looking at the person and item separation indices and reliability estimates. Here, the person separation index is 1.62 and the person reliability is 0.72, both of which are too low to the meet the criteria—this has gotten worse compared to the Round 1 results. However, the item results have gotten better with an item separation index of 2.79 and an item reliability estimate of 0.89. While still a little low, these are much closer to the recommended criteria than in Round 1. Finally, when comparing the ordering of the items in the variable map to the expected ordering in *Table 14*, we notice some disordering of the items. Specifically, Larissa (A1_13) is easier than theorized—consistent with what we saw in Round 1. Different from Round 1, David (A1_7) has a much higher estimate than expected—by approximately two levels, and Viktoria (A1_3) has a slightly higher estimate than expected—by about one level.

Next, we look at the category characteristic curves (CCCs) shown in *Figure 17*, on the following page, to determine whether the response options function as intended. While we can see that the response options are ordered and each response option is dominant in one portion of the distribution, the curve for the fourth response option is extremely flat. More work is needed to improve the use and functioning of the response options.

Figure 17

Comparative Math Anxiety (Emotional-Attitudinal) CCCs, R2 (Expected)

```
        CATEGORY PROBABILITIES: MODES - Andrich thresholds at intersections
P      -+---------+---------+---------+---------+---------+---------+-
R  1.0 +                                                             +
O      |                                                             |
B      |                                                             |
A      |                                                         5555|
B   .8 +11                                                     55    +
I      |  111                                                55      |
L      |     11                                            55        |
I      |       11                                        55          |
T   .6 +         1                                     55            +
Y      |          11                                  5              |
    .5 +            11                               5               +
O      |              12222222222                  55                |
F   .4 +            22211        22 3333333       5                  +
       |         222     11      33*2      333  55                   |
R      |      222          1   33    22     44**44444444             |
E      |   222              **3        22444  5 33      4444         |
S   .2 +222               33  11       4422 55    333       4444     +
P      |               333      111 444   5*22       33         4444 |
O      |           3333         444*11 555    22       3333         4|
N      |     333333       44444  5555*1111     222222     333333     |
S   .0 +*******************5555555         11111111111***************+
E      -+---------+---------+---------+---------+---------+---------+-
       -3        -2        -1         0         1         2         3
        PERSON [MINUS] ITEM MEASURE
```

After looking at the variable map and the CCCs, we next want to review the item and person misfit tables (*Tables 15* and *16*, respectively, on the following pages).

Table 15

Comparative Math Anxiety (Emotional-Attitudinal) Item Fit Statistics, R2 (Expected)

Item	Measure	INFIT		OUTFIT	
		MNSQ	*ZSTD*	*MNSQ*	*ZSTD*
A1_4	-0.58	1.36	2.2	1.50	2.9
A1_9	0.76	1.30	1.7	1.17	0.9
A1_1	0.17	1.12	0.8	1.23	1.4
A1_5	0.73	1.20	1.2	1.09	0.5
A1_11	-0.72	1.12	0.9	1.16	1.1
A1_13	0.07	1.12	0.8	1.10	0.6
A1_2	0.19	1.00	0.1	1.08	0.5
A1_10	0.07	1.02	0.2	0.99	0.0
A1_6	-0.16	0.74	-1.8	0.71	-1.9
A1_12	-0.26	0.66	-2.4	0.74	-1.8
A1_3	-0.29	0.72	-2.0	0.73	-1.9
A1_7	0.02	0.73	-1.8	0.73	-1.8

From *Table 15*, above, we see that three items (A1_4, A1_9, and A1_1) have an INFIT and/or OUFIT MNSQ greater than 1.2 and only item A1_4 has OUTFIT MNSQ greater than 1.4. Notice again A1_4 is one of the two easiest items and A1_9 and A1_1 are two of the hardest items on the emotional-attitudinal mathematics anxiety scale. Using the looser criteria of 1.4, only item A1_4 would have been flagged—again, as one of the easiest items, this is not unexpected. From *Table 16*, on the following page, we see twenty-one students had INFIT and/or OUTFIT MNSQ values greater than 1.4, and nine students had INFIT and/or OUTFIT MNSQ values greater than 1.2 but less than 1.4.

Table 16

Comparative Math Anxiety (Emotional-Attitudinal) Person Fit Statistics, R2 (Expected)

Person	Measure	INFIT		OUTFIT	
		MNSQ	*ZSTD*	*MNSQ*	*ZSTD*
104	-1.68	4.18	4.3	3.98	4.1
46	-0.16	3.02	3.7	2.99	3.6
79	-1.23	2.69	2.9	2.79	3.1
98	-1.52	1.92	1.8	2.27	2.3
25	-1.52	2.19	2.2	2.09	2.1
44	-1.10	2.12	2.2	1.82	1.8
99	-1.68	1.71	1.5	2.02	2.0
80	-2.28	1.98	1.7	1.58	1.1
8	-0.64	1.87	1.9	1.97	2.0
95	-1.86	1.92	1.8	1.65	1.4
41	-0.98	1.89	1.9	1.73	1.6
103	-0.98	1.55	1.3	1.85	1.8
92	-0.75	1.70	1.6	1.78	1.7
42	-0.86	1.61	1.4	1.71	1.6
20	-1.86	1.52	1.2	1.58	1.2
18	-2.05	1.57	1.2	1.26	0.7
93	-1.37	1.57	1.3	1.55	1.3
26	-2.87	1.55	1.0	1.34	0.7
56	-2.54	1.51	1.0	1.10	0.4
96	-2.54	1.31	0.7	1.42	0.9
85	-2.28	1.38	0.9	1.40	0.9
2	-0.98	1.29	0.8	1.39	1.0
21	-1.52	1.39	1.0	1.25	0.7
27	-2.05	1.31	0.8	1.39	0.9
50	-0.75	1.34	0.9	1.36	0.9
90	-2.05	1.33	0.8	1.28	0.7
66	-0.54	1.30	0.8	1.31	0.8
33	-0.54	1.24	0.7	1.21	0.6
32	-0.75	1.21	0.6	1.23	0.7
70	-2.28	1.21	0.6	0.95	0.1

As with the first round of data collection, we want to look at the most unexpected responses using *Table 17*, on the following page, to understand why the misfit may be occurring. Note that because item A1_8 was removed from the scale, items 9-13 are in item locations 8-12, respectively.

Table 17

Comparative Math Anxiety (Emotional-Attitudinal) Poorly Fitting Persons, R2 (Expected)

Person	Response Statistic	Item											
		1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*
104	Obs	1	5	1	1	1	1	1	1	1	5	1	1
	Exp	1.6	1.6	1.8	2	1.4	1.7	1.6	1.3	1.6	2.1	1.8	1.6
	Z		5								3		
46	Obs	1	1	5	5	1	5	1	1	5	5	2	1
	Exp	2.6	2.6	3	3.3	2.1	2.9	2.7	2.1	2.7	3.4	3	2.7
	Z						2			2			
79	Obs	1	1	1	1	1	1	3	3	4	1	2	4
	Exp	1.8	1.8	2.1	2.3	1.5	2.0	1.9	1.5	1.8	2.4	2.1	1.8
	Z								2	2			2
98	Obs	2	2	1	1	3	1	2	3	1	1	1	3
	Exp	1.6	1.6	1.9	2.1	1.4	1.8	1.7	1.4	1.7	2.2	1.9	1.7
	Z					2			2				
25	Obs	2	2	4	2	1	1	1	1	1	1	1	4
	Exp	1.6	1.6	1.9	2.1	1.4	1.8	1.7	1.4	1.7	2.2	1.9	1.7
	Z			2									3

Note. Obs = observed response, Exp = expected response, and Z = z-residual

In contrast to the last round of data collection, we are not seeing as consistent of a pattern such that the students with the greatest misfit have significant misfit (z-residuals of 2 or greater) on the hardest and easiest items. This is because those students have already been removed from this sample as they make up the reversed group of respondents. Here, the misfit appears to be a lot more random with very little to no patterns apparent in the data. Student 104 has significant misfit on items A1_2 and A1_11, student 46 on items A1_6 and A1_10, student 79 on items A1_9, A1_10, and A1_13, etc. Some of the misfit is still occurring because of highly unexpected responses. For example, student 104 was expected to give a response of approximately 2 or "less anxious than" for both items A1_2 and A1_11 but instead responded 5 or "much more anxious than" on both items. Looking across the ten students with the most misfit, we see zero students with significant misfit on item A1_1, three students with significant misfit on item A1_4, and three students with significant misfit on item A1_9 (item location 8). Looking back at *Table 15*, none of the items had INFIT MNSQ greater than 1.4, which looks at the internal consistency of responses, but item A1_4 had OUTFIT MNSQ of 1.5, which is more sensitive to outlier, or unexpectedly extreme, responses. For all three of the ten students that had significant misfit on item A1_4 (students 44, 80, and 95), they were expected to respond approximately 2 or "less anxious than" but selected 4 "much more anxious" or 5 "much more anxious than."

Mental-cognitive (expected) results. The expected variable map for the mental-cognitive scale from the second round of data collection can be found in *Figure 18*, on the following page. Similar to the emotional-attitudinal scale, the persons appear to be fairly well spread but not totally reaching the high end of the continuum. Looking at the items, we see that there is very little spread of the items along the continuum, rather they are all clustering around the mean. These observations are echoed in the statistics—the person separation index is 1.87 and the person reliability is 0.78 both of which are slightly too low to meet the criteria; in addition, the item separation index is 2.21 and the item reliability estimate is 0.83, both of which are too low to meet the criteria (but at least better than they were in Round 1). Finally, when comparing the ordering of the items in the variable map to the expected ordering in *Table 14*, we notice quite a bit more disordering with Jaime (A2_6) having a much lower difficulty than expected, Emilie (A2_9) and Victor (A2_1) having a slightly lower difficulty than expected, and item A2_7 (Lena) having a much higher difficulty than expected. As one of the theorized easiest items, it is particularly surprising to see the disordering from item A2_7 (Lena).

Figure 18

Comparative Math Anxiety (Mental-Cognitive) Variable Map, R2 (Expected)

```
MEASURE    PERSON - MAP - ITEM
               <more>|<rare>
    2                +
                     |
                     |
                     |
                     |
                     |
                     |
                     |
                  . T|
                 .#  |
    1                +
                  .  |
                 .#  |
                 ##  |
                .##  |
                .##  |
              .#### S|
              #####  |T
           .#######  |  A2_9_Emilie
             ######  |S A2_10_Joseph  A2_1_Victor   A2_2_Ana
    0   ###########  +M A2_5_Petra    A2_6_Jaime    A2_7_Lena     A2_8_Leon
           .#######  |S
           .#######  |  A2_11_Lucas
          ######### M|T A2_3_Carlos
            .######  |
               .###  |
              #####  |
               .###  |
                ###  |
                .##  |
   -1      .#######  +
                    S|
                ###  |
                .##  |
                     |
                  #  |
               .###  |
                     |
                 ##  |
                    T|
   -2                +
                  #  |
                     |
                     |
                  .  |
                     |
                     |
                     |
                  #  |
                     |
   -3                +
                     |
                     |
                     |
                  #  |
                     |
                     |
                     |
                     |
                     |
   -4        .#####  +
               <less>|<freq>
 EACH "#" IS 3: EACH "." IS 1 TO 2
```

Next, we look at the category characteristic curves (CCCs) shown in *Figure 19* to determine whether the response options function as intended. Overall, the curves look pretty good, with the response options ordering as they should. However, we would still like to see the second and fourth response options dominating a bit more in their portions of the distribution.

Figure 19

Comparative Math Anxiety (Mental-Cognitive) CCCs, R2 (Expected)

```
        CATEGORY PROBABILITIES: MODES - Andrich thresholds at intersections
P      -+---------+---------+---------+---------+---------+---------+-
R  1.0 +                                                             +
O      |                                                             |
B      |                                                             |
A      |111                                                      5555|
B   .8 +   111                                                 55    +
I      |      11                                            555      |
L      |        11                                        55         |
I      |          11                                     5           |
T   .6 +            1                                  55            +
Y      |             11                               5              |
    .5 +               1                            55               +
O      |                11                         5                 |
F   .4 +                  1       3333333333     55                  +
       |              22222**22**3          3344*444                 |
R      |           222       *3  222      44443*    44444            |
E      |        222        33 11    222 44   55 33       444         |
S   .2 +    2222         33     11   44*2  55     33        4444     +
P      |2222           33         1*4    **         333         4444 |
O      |           3333        4444 11 55  222         333          4|
N      |    3333333       44444   5555*1111   222222      3333333    |
S   .0 +******************55555555         111111111*****************+
E      -+---------+---------+---------+---------+---------+---------+-
       -3        -2        -1         0         1         2         3
        PERSON [MINUS] ITEM MEASURE
```

Finally, we look at the item and person misfit tables (*Tables 18* and *19*, respectively). From *Table 18*, on the following page, we see that only one item (A2_7) has an INFIT and OUTFIT MNSQ greater than 1.2 (but less than 1.4). This item is expected to be one of the easiest items but was actually the fourth easiest. Given how disordered this item was in the variable map, it is not surprising that there are some misfit issues.

Table 18

Comparative Math Anxiety (Mental-Cognitive) Item Fit Statistics, R2 (Expected)

Item	Measure	INFIT		OUTFIT	
		MNSQ	*ZSTD*	*MNSQ*	*ZSTD*
A2_7	0.00	1.26	3.4	1.28	3.6
A2_3	-0.31	1.18	2.5	1.19	2.6
A2_9	0.17	1.08	1.1	1.02	0.3
A2_1	0.12	1.06	0.8	1.06	0.9
A2_6	0.02	1.04	0.6	1.00	0.0
A2_11	-0.18	0.99	-0.1	1.03	0.4
A2_2	2:52	0.91	-1.3	0.91	-1.2
A2_8	0.02	0.87	-1.9	0.85	-2.1
A2_10	0.08	0.87	-2.0	0.84	-2.3
A2_5	-0.05	0.80	-3.0	0.78	-3.2

From *Table A1*, in Appendix A, we see seventy-nine students had INFIT and/or OUTFIT MNSQ values greater than 1.4, and an additional thirty-four students had INFIT and/or OUTFIT MNSQ values greater than 1.2 but less than 1.4.

Again, we can look at *Table 19*, on the following page, to see the observed and expected responses for the students with the greatest misfit. Here, item A2_4 was deleted from the scale so items 5-11 are in item locations 4-10, respectively. As with the emotional-attitudinal scale, there are not really any consistent patterns to the misfit although the misfit does appear to be because of highly unexpected responses. For example, responding 5 or “much more anxious than” when a student was expected to respond with a 1 “much less anxious than” or a 2 “less anxious than.” Several of the students, i.e., students 313, 28, and 10, have misfit that is more a matter of degree. That is, they responded 1 “much less anxious than” or 5 “much more anxious than” when they were approximately expected to response 3 or “as anxious as.”

Table 19

Comparative Math Anxiety (Mental-Cognitive) Poorly Fitting Persons, R2 (Expected)

Person	Response Statistic	Item									
		1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*
198	Obs	1	1	1	1	1	1	1	1	1	5
	Exp	1.4	1.4	1.5	1.4	1.4	1.4	1.4	1.3	1.4	1.5
	Z										5
306	Obs	1	1	5	1	1	5	1	1	1	1
	Exp	1.7	1.7	2	1.8	1.8	1.8	1.8	1.7	1.7	1.9
	Z			3			3				
322	Obs	1	1	5	1	1	5	1	1	1	1
	Exp	1.7	1.7	2	1.8	1.8	1.8	1.8	1.7	1.7	1.9
	Z			3			3				
313	Obs	5	4	5	1	2	5	1	1	4	1
	Exp	2.8	2.8	3.2	2.9	2.9	2.9	2.9	2.7	2.8	3.1
	Z	2					2				-2
220	Obs	5	1	2	1	5	3	2	5	1	5
	Exp	2.9	2.9	3.3	3.0	3.0	3.0	3.0	2.8	2.9	3.2
	Z	2			-2	2			2		

Note. Obs = observed response, Exp = expected response, and Z = z-residual

Only one item had INFIT and OUTFIT MNSQ greater than 1.2 according to *Table 18*—item A2_7. At item location 6, this item showed significant misfit (z-residual greater than the absolute value of 2) when students responded 5 or "much more anxious than" when they were expected to respond 2 "less anxious than" or 3 "as anxious as." As an item that was theorized to be one of the two easiest items but ended up in the middle of the item difficulty range, this extreme misfit provides some evidence as to why this item may have not ordered as expected.

Physical-somatic (expected) results. The expected variable map for the physical-somatic scale from the second round of data collection can be found in *Figure 20*, on the following page. Notice that the persons and items both actually appear to be fairly well spread, but similar to the other two dimensions, the persons are trending towards the lower half of the continuum while the items trend (only a little in this case) towards the upper half.

Figure 20

Comparative Math Anxiety (Physical-Somatic) Variable Map, R2 (Expected)

```
MEASURE PERSON - MAP - ITEM
           <more>|<rare>
    1          T+
             XX  |  A3_11_Martin   A3_3_Laila
                 |S A3_12_Liam
              X  |  A3_6_Jasper
              X  |
             XX  |  A3_5_Angela
    0       XXX  +M A3_10_Erica    A3_1_Ivan      A3_2_Gabriell  A3_9_Noah
           XXXX S|
            XXX  |  A3_7_Nora
      XXXXXXXXX  |
              X  |S
          XXXXX  |
   -1         X  +
              X  |  A3_8_Romita
           XXXX M|T A3_4_Jay
          XXXXX  |
           XXXX  |
              X  |
   -2      XXXX  +
             XX  |
            XXX S|
                 |
            XXX  |
          XXXXX  |
   -3            +
              X  |
                 |
                T|
              X  |
                 |
   -4            +
                 |
              X  |
                 |
                 |
                 |
   -5       XXX  +
           <less>|<freq>
```

Looking at the separation indices and reliability estimates, we see that the person separation index is 2.14, the person reliability is 0.82, the item separation index is 3.97, and the item reliability estimate is 0.94, all of which are acceptable and meet the minimum criteria. This is the first dimension to meet these criteria in the scale development process so far. Finally, when comparing the variable map item ordering to the expected ordering in *Table 14*, we should note that there is still quite a bit of disordering with Noah (A3_9) slightly more difficult than theorized (similar to Round 1); Erica (A3_10) and Ivan (A3_1) slightly easier than expected, and Liam (A3_12) way more difficult than expected. So while we have adequate spread of the items, more work is needed to clarify the construct.

Next, we look at the category characteristic curves (CCCs) shown in *Figure 21*. While we wouldn't mind seeing the middle response options (2, 3, and 4) being a little more dominant in their respective area of the distribution, these curves look very much how we would like, meaning the response options are being utilized as intended.

Figure 21

Comparative Math Anxiety (Physical-Somatic) CCCs, R2 (Expected)

```
        CATEGORY PROBABILITIES: MODES - Andrich thresholds at intersections
P      -+---------+---------+---------+---------+---------+---------+-
R  1.0 +                                                             +
O      |                                                             |
B      |                                                             |
A      |                                                             |
B   .8 +1                                                           5+
I      | 111                                                     55 |
L      |    11                                                555   |
I      |      11                                             5      |
T   .6 +        1                                          55       +
Y      |         11                                      55         |
    .5 +           1             3333333333              5           +
O      |            11         33          333         55            |
F   .4 +              *22222 33               33444444*4             +
       |          2222 11  3*222             44433  55  4444         |
R      |      2222       *3     22        444     3*        444      |
E      |   222         33 11      222   44       55 33         444   |
S   .2 +222          33     11       ***       55     33          444+
P      |           33         11  444   22   55         33           |
O      |       3333            4**1       ***2            3333       |
N      | 333333          444444    1*****5    222222          333333 |
S   .0 +*****************5555555555      11111111111*****************+
E      -+---------+---------+---------+---------+---------+---------+-
       -3        -2        -1         0         1         2         3
        PERSON [MINUS] ITEM MEASURE
```

Finally, *Tables 20* and *21* show the person and item misfit statistics, respectively. From *Table 20*, below, we see that five items (A3_3, A3_8, A3_11, A3_4, and A3_12) have an INFIT or OUFIT MNSQ greater than 1.2 but less than 1.4.

Table 20

Comparative Math Anxiety (Physical-Somatic) Item Fit Statistics, R2 (Expected)

Item	Measure	INFIT		OUTFIT	
		MNSQ	*ZSTD*	*MNSQ*	*ZSTD*
A3_3	0.78	1.18	1.0	1.27	1.1
A3_8	-1.19	1.25	1.5	1.24	1.4
A3_11	0.87	1.24	1.3	0.97	0
A3_4	-1.40	1.22	1.4	1.21	1.3
A3_12	0.75	1.21	1.2	1.02	0.2
A3_5	0.14	1.08	0.5	0.99	0
A3_7	-0.38	0.99	0.0	1.05	0.4
A3_10	-0.05	0.96	-0.2	0.98	0
A3_2	-0.01	0.97	-0.1	0.88	-0.6
A3_1	0.02	0.76	-1.5	0.8	-1
A3_6	0.49	0.72	-1.7	0.64	-1.8
A3_9	-0.01	0.66	-2.2	0.67	-1.9

This is an improvement from Round 1, where two items had an INFIT MNSQ greater than 2. As expected, these five items are the easiest (A3_8 and A3_4) and hardest (A3_3, A3_11, and A3_12) on the physical-somatic mathematics anxiety scale. As described previously, some misfit at the ends of the scale is not unexpected. From *Table 21*, on the following page, we see thirteen students had INFIT and/or OUTFIT MNSQ values greater than 1.4, and three students had INFIT and/or OUTFIT MNSQ values greater than 1.2 but less than 1.4.

Table 21

Comparative Math Anxiety (Physical-Somatic) Person Fit Statistics, R2 (Expected)

Person	Measure	INFIT		OUTFIT	
		MNSQ	*ZSTD*	*MNSQ*	*ZSTD*
11	-0.12	3.28	3.9	3.29	3.9
44	-0.46	3.11	3.8	3.14	3.8
25	-2.58	1.94	1.6	3.02	2.5
18	0.34	2.80	3.2	2.78	3.2
9	-0.57	2.61	3.2	2.63	3.2
21	-1.28	2.45	3.0	2.41	2.9
31	-0.12	2.26	2.5	2.26	2.5
3	-1.68	1.92	2.0	2.13	2.3
28	-1.83	1.79	1.7	1.88	1.8
62	-1.68	1.73	1.7	1.58	1.4
13	-0.91	1.58	1.5	1.66	1.7
68	-2.84	1.66	1.2	1.52	0.9
37	-1.99	1.42	1.0	1.32	0.8
39	-1.28	1.25	0.8	1.31	0.9
42	-0.80	1.25	0.8	1.26	0.8
60	-1.99	1.21	0.6	1.07	0.3

Looking at *Table 22*, on the following page, with the observed and expected responses for the students with the greatest misfit, we see that there don't really appear to be patterns in the misfit. Given that there isn't a lot of significant misfit in the items, that is all INFIT and OUTFIT MNSQs are less than 1.4, this is promising and makes some sense that there wouldn't be visible patterns in the misfit.

Table 22

Comparative Math Anxiety (Physical-Somatic) Poorly Fitting Persons, R2 (Expected)

Person	Response Statistic	Item											
		1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*
11	Obs	3	4	3	5	1	3	1	1	4	5	4	1
	Exp	2.9	2.9	2.3	3.9	2.8	2.6	3.2	3.8	2.9	3.0	2.3	2.4
	Z					-2		-2	-3		2		
44	Obs	4	5	1	4	2	1	3	1	2	3	5	1
	Exp	2.7	2.7	2.1	3.7	2.6	2.3	3.0	3.5	2.7	2.7	2.0	2.1
	Z		2						-2			3	
25	Obs	1	2	3	1	1	1	1	1	1	1	2	2
	Exp	1.3	1.4	1.2	2.1	1.3	1.2	1.5	2.0	1.4	1.4	1.1	1.2
	Z			4								2	2
18	Obs	5	3	2	3	1	3	5	5	5	5	1	1
	Exp	3.2	3.3	2.7	4.2	3.1	2.9	3.5	4.1	3.3	3.3	2.6	2.7
	Z	2				-2				2			
9	Obs	2	1	3	4	1	3	4	2	1	2	3	5
	Exp	2.6	2.6	2	3.6	2.5	2.2	2.9	3.5	2.6	2.6	1.9	2.0
	Z												3

Note. Obs = observed response, Exp = expected response, and Z = z-residual

A Deeper Dive: Student Cognitive Interviews

Students' perception of the survey. While the results of the cluster and Rasch analyses provide some very solid evidence to support the hypothesis that some students were matching the response options to the items (reversed) and some students were comparing themselves to the items using the response options (expected), cognitive interviews were conducted with students in fall 2019 to collect additional data on the phenomenon. In total, five 9th grade students who were similar to the Round 2 sample in ability level completed the survey and participated in a cognitive interview.

Overall, two students could be represented as falling in the reversed group and three students as falling in the expected group. When asked how they interpreted the directions or what the directions asked them to do, one student that would have been in the reversed group said "[the section asked me to] talk about how I would feel based on how they could feel, kind of, so like when he feels stressed out, that kind of made me feel stressed out too… I was kind of thinking about if they were like here with me, how would I feel like just with them being around and like in that state of mind." The other student in the reversed group similarly said "well for this [item] I thought like completely on edge, like how I feel when I feel completely on edge is do I feel: is that less anxious, much less anxious, more anxious? I was thinking like which [item] is which [level of] stress." These quotes from students directly support the hypothesis presented.

For the three students that responded as expected, the general sense of the directions, as stated by one of the students, was to "talk about how you feel compared to how it [the person in the sentence] says they feel." While three students did respond as expected, they did have some input on the survey. Overall, they felt that the survey was pretty long—part of the hypothesis for why students may not have used the response options as intended. In addition, while they were able to figure out the directions eventually, one student did note their initial confusion with the question format and that "it took me a couple of questions to completely understand [the directions]." When presented as an option, they agreed that it would make more sense to be asked to select the statement(s) that sounded similar to themselves and their experience. Finally, while the third student typically responded as expected mostly selecting 'less anxious' and 'much less anxious' as appropriate, they also selected 'less anxious' for the item 'Daniel feels in his element'; when probed, the student admitted they should have put "the same, like as anxious" given their similarly low (or non-existent) level of anxiety. This supports the data presented in many of the item misfit tables, particularly with items that fall into the lowest level of anxiety.

In summary, the cognitive interviews conducted with 9th grade students who took the same survey as those students in the Round 2 sample provided ample evidence to confirm the hypothesis that some students interpreted the directions incorrectly, matching items to response options rather than comparing themselves to the items using the response options. Additional information given by the students provided support for other aspects of the results as well.

Students' perception of the items and their relationship to the construct. As described in Chapter 3, a second set of cognitive interviews was also conducted. In this set of interviews, students were asked to participate in a type of card-sorting (or item-sorting activity), where they were provided the set of items from one dimension of the construct on individual slips of paper and asked to sort them along a continuum from low to high mathematics anxiety. Questions were asked during and after the activity based on how they ordered the items, if they grouped any items, and if they seemed to struggle with the task. In all, ten students participated in this set of cognitive interviews. Major takeaways are presented below.

The first major takeaway was that the items were interpreted as being true to students' experiences, particularly for those that had a strong sense of the construct or had experienced mathematics anxiety themselves. One student said it was "encouraging that they [the items] all seem really accurate to the way students feel," while another felt the items all seemed to capture some of their experience in math at varying points [in time]. This particular students' response brought up another interesting and important point—that the construct of mathematics anxiety is fluid and can change over time. Therefore, it is important to have items that can capture that breadth of experience and any potential change in that experience.

The second major takeaway was that many of the items felt very similar to the students and they had trouble separating some of the items, instead choosing to group them or place them at the same level rather than trying to order them in a ladder-like progression. Some examples of phrases that were deemed similar are: blanking out and mind completely freezes; un-phased and in their element; and a little intimidated and a bit apprehensive. While it may be important in the item-development phase to generate as many items as possible and redundancy is appropriate, as we move towards finalizing the scale, it is appropriate and necessary to remove some of the redundancy. Students' perceptions of where that redundancy lies, supported by the item locations, or item difficulties, from prior survey data collection, provided data and evidence to support those decisions.

Final Revisions: Round 3

After analyzing the Round 2 and cognitive interview data, it was now time to make the final set of instrument revisions for the study. For each dimension of mathematics anxiety, I looked across the Round 2 results (including the initial, reversed, and expected results) as well as the results of the cognitive interviews item-sorting activity. As needed, the Round 1 results were also referenced in regards to the consistency of items or how changes across versions impacted the results. As noted previously, each dimension of the scale has a decent number of items and some of those items have been inconsistent across administrations. It was important during the survey development process to generate as many items as possible; however, in this final round of revisions, I took an eye at reducing the number of items if I believed it would result in increased clarity and consistency while still maintaining construct representation across the continuum from low to high mathematics anxiety. By reducing the number of items and the item redundancy, it was hypothesized that it would be easier for students to see differences across the items and that the items would be more likely to stabilize in their theorized locations. Specific revisions for each of the three dimensions of the Comparative Mathematics Anxiety Scale are described next and can be found summarized in *Table 23*, on the following page. Note that deleted items have been stricken-through and items that have moved levels from the second round to the third round have been italicized in the table.

Table 23

Comparative Mathematics Anxiety Items, Revisions Round 3 (R3, Final)

Level of Anxiety	Dimension of Mathematics Anxiety		
	Emotional/Attitudinal	*Mental/Cognitive*	*Physical/Somatic*
Overwrought	1_5. Ruby is terrified. ~~1_9. Isaac feels intense fear.~~	~~2_6. Jaime thinks about how anxious math makes him feel instead of the problem in front of him.~~ 2_9. Emilie can think about nothing else except her fear of math.	3_3. Laila gets nauseous to the point of feeling like she is going to be sick. ~~3_11. Martin is on the verge of tears.~~
Distressed	~~1_1. Michael is completely on edge.~~ ~~1_2. Sebastian feels extremely overwhelmed.~~ 1_10. Andrea is distressed. *1_13. Larissa feels emotionally worked up.*	~~2_1. Victor's mind completely freezes.~~ *2_10. Joseph's mind keeps blanking out.*	*3_2. Gabrielle's heart is racing.* ~~3_6. Jasper gets a stomachache.~~
Antsy	~~1_6. Alicia feels uneasy.~~ 1_12. Gabriel has a sense of worry.	2_2. Ana can't recall what she learned.	*3_1. Ivan feels physically uncomfortable.* *3_9. Noah's palms start to sweat.* *3_10. Erica gets a headache.* ~~3_5. Angela bites her nails.~~ ~~3_12. Liam cannot stop squirming in his seat.~~
Apprehensive	1_3. Viktoria is a bit apprehensive. [moved to between A1_10 & A1_11] 1_7. David feels a little intimidated.	2_5. Petra feels confused. 2_8. Leon's thoughts are scattered. ~~2_11. Lucas feels like he should know how to solve the problem but he can't put his thoughts in order.~~	3_7. Nora gets a little tense.
Tranquil	1_4. Daniel feels in his element. 1_11. Chloe feels unfazed.	2_3. Carlos is able to think clearly. 2_7. Lena is focused on the problems in front of her.	3_4. Jay is completely calm. 3_8. Romita feels completely at ease.

Emotional-attitudinal anxiety. To begin the revision process, I started with the items at the extreme ends of the scale. Items A1_5 and A1_9 have both been consistently at the top, but to reduce redundancy item A1_9 was removed from the scale. Similarly, items A1_11 and A1_4 have both been consistently at the bottom, but rather than deleting one of the items, both were kept to help maintain some balance of positive and negative items. Items A1_13 (worked up) and A1_1 (on edge) were both consistently moderately-high, but Item A1_1 was chosen for removal to reduce redundancy. It was also decided, from the theoretical perspective, to move item A1_13 (worked up) from the highest level of mathematics anxiety (overwrought) to the second highest level (distressed). Item A1_2 (extremely overwhelmed) was inconsistent across administrations and deleted. Item A1_6 (uneasy) was also removed to reduce redundancy at the middle level as it was more inconsistent than the other item theorized to be at the middle level—A1_12 (worry). As a final note, item A1_3 (apprehensive) was moved to be between items A1_10 and A1_11 for survey administration so that it was not the first item from the dimension students would see when taking the survey, since items A1_1 and A1_2 were both removed; however, for analysis, the items were re-ordered so that item A1_3 makes up the first data column.

Mental-cognitive anxiety. At the high end of the mental-cognitive dimension, it was decided to keep item A2_9 over A2_6, as item A2_6 was inconsistent across administrations. Following the same reasoning as for the emotional-attitudinal dimension, both items at the low end of the scale (A2_3 and A2_7) were kept. Items A2_1 (mind freezes) and A2_10 (blanking out) were both pretty consistently moderately-high; as item A2_10 was more consistently moderately-high, it was kept and moved theoretically from the middle level (antsy) to the second-highest level (distressed). Of the remaining items, items A2_2 (can't recall), A2_5 (confused), and A2_8 (thoughts scattered) were kept in the scale and at their theorized level, while item A2_11 (can’t put thoughts in order) was removed because of inconsistencies across the survey administrations as well as in the item-sorting interviews.

Physical-somatic anxiety. Items A3_3 (nauseous) and A3_11 (tears) were consistently high across administrations but it was decided to delete one of the items to reduce redundancy. It was theorized item A3_11 might show differential item functioning by gender, so it was removed from the scale. As with the other two scales, both of the lowest items—A3_4 and A3_8—were kept. Item A3_2 (heart racing) was consistently moderate to moderately-high across administrations, so the item was kept and moved theoretically from the middle level to the second-highest level. Items A3_6 (stomachache) and A3_10 (headache) are both consistent and somewhat similar, so it was decided to remove one to reduce redundancy. The decision was made to keep item A3_10 as it is a more moderate item and less repetitive content-wise with item A3_3 (nauseous). Then the decision was made to move item A3_10, along with item A3_1 (physically uncomfortable), from the second-highest level to the middle level. Item A3_9 (palms start to sweat) was also moved to the middle level but from the second-lowest level. Finally, items A3_5 (bites nails) and A3_12 (squirms in seat) were both removed because of the determination that the items were too construct-irrelevant.

Once the revisions were finalized, data was collected in fall 2020 from students enrolled in a mathematics course at the same high school where the cognitive interviews were conducted (Step 4). In this round of data collection, the survey included all three dimension of the Comparative Mathematics Anxiety Scale, the validation scales, and a few demographic questions. Because of the COVID-19 pandemic, the survey was administered electronically using the Qualtrics survey platform. The results of this third and final round of data collection are presented next (Step 5) and provide evidence for examining the match between the items and the mathematics anxiety conceptual framework (Step 6).

Final Results: Round 3

Emotional-attitudinal scale. The variable map for the emotional-attitudinal scale from the final round of data collection can be found in *Figure 22*, below.

Figure 22

Comparative Math Anxiety (Emotional-Attitudinal) Variable Map, R3

```
MEASURE   PERSON - MAP - ITEM
              <more>|<rare>
    3               +
                    |
                    |
                 X  |
                    |T
    2           XX T+  A1_5_Ruby
               XXX  |
                 X  |
                    |
              XXXX  |
    1         XXXX  +S
              XXXX  |  A1_10_Andrea
             XXXXX S|
                    |  A1_13_Larissa
        XXXXXXXXXX  |  A1_3_Viktoria
    0       XXXXXX  +M
          XXXXXXXX  |  A1_12_Gabriel  A1_7_David
             XXXXX  |
                 X  |
      XXXXXXXXXXXX M|
   -1       XXXXXX  +S
          XXXXXXXX  |
                 X  |  A1_4_Daniel
            XXXXXX  |  A1_11_Chloe
          XXXXXXXX  |
   -2      XXXXXXX  +
                   S|T
         XXXXXXXXX  |
          XXXXXXXX  |
                    |
   -3         XXXX  +
                 X  |
                 X  |
                   T|
                    |
   -4               +
                    |
                 X  |
                    |
                    |
   -5               +
              <less>|<freq>
```

Overall, the persons and items appear to be very well spread, with a couple of gaps in the item continuum. Looking at the person and item separation indices and reliability estimates, the person reliability estimate is adequate at 0.80 but the separation index is just slightly too low at 1.99. In contrast, the item statistics are great with a separation index of 7.53 and a reliability estimate of 0.98. Comparing the variable map to the construct map (*Table 23*), we see that the items order mostly as expected; however, both of the items hypothesized to be in the second-lowest level are harder than expected—item A1_3 much higher and item A1_7 slightly higher.

Next, we look at the category characteristic curves (CCCs) shown in *Figure 23*, on the following page, to determine whether the response options function as intended. We can see that (1) all five response options were used by respondents, shown visually by the dominance of each response option in one area of the participants' anxiety level distribution; and (2) the response options are ordered, such that the probability of moving from one category increases in the expected order from much less anxious to much more anxious. Overall, the CCCs look great and the response options appear to function as intended.

Figure 23

Comparative Math Anxiety (Emotional-Attitudinal) CCCs, R3

```
        CATEGORY PROBABILITIES: MODES - Andrich thresholds at intersections
P      -+------+------+------+------+------+------+------+------+-
R  1.0 +                                                         +
O      |                                                         |
B      |                                                         |
A      |                                                         |
B   .8 +                                                        5+
I      |11                                                     5 |
L      |  1                                                  55  |
I      |   11                                              5    |
T   .6 +     1                    33333                   55     +
Y      |      1      222        33     333               5       |
    .5 +       11 222   222   33          3    4444444  5        +
O      |        2*         223             3*44       4*         |
F   .4 +      22  1        3322            4 3        5 44       +
       |    22     1      3    22        44   33    55    44     |
R      |  22        11  33       2      4       3  5        44   |
E      | 2            13          22  44         35           44 |
S   .2 +2            3311           *4          5533            4+
P      |           33    11       44 222      55    33           |
O      |         33        11  444      22  55        333        |
N      |   333333         444**111     555**222          33333   |
S   .0 +******************55555555*****11111111******************+
E      -+------+------+------+------+------+------+------+------+-
       -4     -3     -2     -1      0      1      2      3      4
        PERSON [MINUS] ITEM MEASURE
```

After looking at the variable map and the CCCs, we next want to review the item and person misfit tables (*Tables 24* and *25*, respectively). Two items in *Table 24*, on the following page, have INFIT and OUTFIT MNSQ greater than or equal to 1.4—item A1_13 (Larissa) and item A1_4 (Daniel).

Table 24

Comparative Math Anxiety (Emotional-Attitudinal) Item Fit Statistics, R3

Item	Measure	INFIT		OUTFIT	
		MNSQ	*ZSTD*	*MNSQ*	*ZSTD*
A1_13	0.50	1.46	3.3	1.46	3.4
A1_4	-1.47	1.40	2.9	1.44	3.2
A1_11	-1.54	1.30	2.3	1.35	2.6
A1_5	1.99	1.13	1.0	1.06	0.5
A1_10	0.72	0.81	-1.6	0.82	-1.5
A1_12	-0.23	0.73	-2.4	0.73	-2.4
A1_3	0.15	0.61	-3.6	0.61	-3.7
A1_7	-0.12	0.55	-4.4	0.56	-4.3

In addition, item A1_11 (Chloe) has an INFIT and OUTFIT MNSQ greater than 1.2. While this is not unexpected for items A1_4 and A1_11 as they are the two easiest items, more exploration is needed to understand the misfit problems for item A1_13. Looking at *Table 25*, on the following page, we see that 25 out of 126 persons (approx. 20%) had INFIT and/or OUTFIT MNSQ greater than or equal to 1.4, and an additional eight students had an INFIT and OUTFIT MNSQ greater than or equal to 1.2.

Table 25

Comparative Math Anxiety (Emotional-Attitudinal) Person Fit Statistics, R3

Person	Measure	INFIT		OUTFIT	
		MNSQ	*ZSTD*	*MNSQ*	*ZSTD*
15	0.52	8.13	6.4	8.00	6.3
18	1.52	5.69	5.2	6.63	5.8
60	1.78	4.93	4.6	5.86	5.3
121	0.52	5.06	4.6	5.01	4.5
88	-2.65	3.84	3.7	4.22	3.7
48	0.27	2.86	2.7	2.92	2.8
41	-3.26	1.94	1.5	2.91	2.1
62	-1.50	2.04	1.8	2.41	2.3
113	-0.98	2.26	2.1	2.21	2.0
110	0.9	2.14	1.8	2.10	1.8
89	0.02	2.01	1.8	1.97	1.7
107	-0.98	1.85	1.6	1.81	1.5
64	0.77	1.78	1.5	1.79	1.5
116	-1.24	1.79	1.5	1.78	1.5
83	-0.98	1.70	1.4	1.68	1.3
117	0.27	1.69	1.3	1.64	1.3
25	1.27	1.62	1.3	1.60	1.2
28	-2.34	1.59	1.2	1.43	0.9
118	0.02	1.55	1.1	1.56	1.1
61	-0.23	1.51	1.1	1.53	1.1
93	0.77	1.50	1.1	1.51	1.1
4	-2.65	1.32	0.8	1.48	1.0
12	-0.48	1.41	0.9	1.38	0.9
90	0.27	1.40	0.9	1.37	0.8
137	1.02	1.40	0.9	1.40	0.9
38	-1.77	1.36	0.8	1.39	0.9
46	-1.50	1.37	0.8	1.31	0.8
102	-0.73	1.35	0.8	1.34	0.8
24	0.27	1.30	0.7	1.31	0.7
2	-2.34	1.25	0.7	1.17	0.5
39	-2.99	1.25	0.6	1.12	0.4
50	-0.48	1.23	0.6	1.21	0.6
59	-2.34	1.21	0.6	1.20	0.5

To understand more why the misfit is occurring, we can look at the students with the greatest misfit in *Table 26*, below. Note that because several items were deleted, the items (and their item order locations) are as follows: A1_3 (location 1), A1_4 (location 2), A1_5 (location 3), and A1_7 (location 4), with items A1_10 through A1_13 in locations 5-8.

Table 26

Comparative Math Anxiety (Emotional-Attitudinal) Poorly Fitting Persons, R3

Person	Response Statistic	Item							
		1	*2*	*3*	*4*	*5*	*6*	*7*	*8*
15	Obs	3	1	5	3	3	1	5	5
	Exp	3.2	4.0	2.3	3.3	2.9	4.0	3.4	3.0
	Z		-4	3			-4	2	2
18	Obs	4	1	5	4	4	3	4	5
	Exp	3.7	4.5	2.8	3.8	3.4	4.5	3.9	3.5
	Z		-5	3			-2		2
60	Obs	4	3	5	3	5	2	4	5
	Exp	3.8	4.6	2.9	4.0	3.5	4.6	4.0	3.6
	Z		-2	3		2	-4		
121	Obs	4	1	3	4	4	2	3	5
	Exp	3.2	4.0	2.3	3.3	2.9	4.0	3.4	3.0
	Z		-4				-2		2
88	Obs	2	1	2	1	1	1	2	4
	Exp	1.6	2.4	1.1	1.7	1.4	2.4	1.8	1.5
	Z			2			-2		4

Note. Obs = observed response, Exp = expected response, and Z = z-residual

Items A1_4 (location 2) and A1_11 (location 6) are the easiest items and were both flagged for misfit. From *Table 26*, we can see that many individuals have misfit on these items and typically in the patterns we were seeing in the first round of data collection where students were expected to respond 4 "more anxious than" or 5 "much more anxious than" but instead responded with a 1 "much less anxious than," a 2 "less anxious than," or a 3 "as anxious as." So, it appears that with the easiest items there are some individuals who are matching the items to response options rather than comparing themselves to the items using the response options.

There was also some misfit for item A1_13 in item location 8. Here, it appears we are seeing significant misfit (z-residual of 2 or greater) when students answered 5 “much more anxious than” or 4 “more anxious than” but were expected to give a response of 3 “as anxious as” or 2 “less anxious than.” Item A1_13 reads “Larissa feels emotionally worked up.” It is possible that students are interpreting this question different, with most seeing the item as a relatively high anxiety item (hence its relatively high item difficulty location), but some students seeing the items as a lower-anxiety item and thus are responding that they are “more anxious than” or “much more anxious than” Larissa. Without additional data, not much else can be determined about this item. However, it is one that should be considered for potential future revision or elimination from the scale.

Given this is the final round of data collection, we also ran an analysis of the Rasch residuals to determine whether or not all, or at least most, of the common variability among the items has been explained by the construct or whether more is left to be explained in the data. To do this, the standardized residuals from the analysis were exported from Winsteps and imported into SPSS, where a principal components analysis (PCA) was then performed on the residuals.

First, we needed to determine the appropriateness of conducting a data reduction analysis, like principal components analysis, on the residual correlation matrix. We do this by looking at the following indicators: the determinant, KMO estimate of sampling adequacy, and Bartlett’s Test of Sphericity. The determinant indicates the amount of variability in the correlation, covariance matrix. A non-zero determinant indicates there is enough unique, uncorrelated variability in the residuals to conduct a factor or principal components analysis; therefore, we want a determinant greater than 0, indicating no perfectly correlated variability in the residuals. Here, the determinant was 0.027, indicating that there is some unique variability left in the residuals. Next, the KMO is an estimate of sampling adequacy that ranges from 0 to 1, where 0 indicates the items do not share a common component and 1 indicates the items share a common component. For the residual analysis, smaller KMO values are preferred. The KMO for this analysis was 0.111 indicating there may be a common component remaining but a relatively weak one. Finally, Bartlett's Test of Sphericity and corresponding significance value is another way of indicating the extent of variation in the item correlation matrix, or in this case, the item residual correlation matrix. A significant Bartlett's test is typically preferred, but a non-significant Bartlett's test is desired in a residual analysis. Unfortunately, Bartlett's Test of Sphericity was highly significant with X^2 (28) = 439.554, $p < .0005$, meaning there is quite a bit of variation left in the item residual correlation matrix. Taken together, these indicators suggest that once the construct variance has been accounted for by the Rasch analysis there is minimal variance remaining in the residuals to support conducting a principal components analysis.

Recognizing the weak support for conducting a principal components analysis, one was still conducted in order to inspect what dependencies might remain between individual pairs of items. Results for all residual analyses can be found in Appendix B. According to the variance table (*Table B1*), three components have an eigenvalue greater than 1; however, looking at the scree plot (*Figure B1*), we see that there is likely only one meaningful component left in the residuals, as there is an appreciable drop in the eigenvalues after the first component. It is important to point out that if the residual correlations were all perfectly zero, then the expected eigenvalue for each item would be 1—that is, no items shared any common variance. Unfortunately, we would rather see a flat line in the screen plot with all eigenvalues equal to 1 in a residual analysis. To determine what variability there might be left to explain in the residuals, we look at the component loading matrix (*Table B2*) and rotated component plot for the first two extracted components (*Figure B2*). In the table, we see that items A1_3 (Viktoria), A1_4 (Daniel), A1_7 (David), and A1_11 (Chloe) have negative component loadings while the other items have positive loadings on the first component. These were theorized to be the four easiest items, although that did not occur in practice with item A1_3 much higher than expected. In the figure, we see the residuals are for the most part randomly scattered along the two components; however, some items are still highly correlated (e.g., z2 and z6 or items A1_4 and A1_11). More work is needed to understand what exactly is left to explain in the items.

Mental-cognitive scale. The variable map for the mental-cognitive scale from the last round of data collection can be found in *Figure 24*, on the following page.

Figure 24

Comparative Math Anxiety (Mental-Cognitive) Variable Map, R3

```
MEASURE    PERSON - MAP - ITEM
               <more>|<rare>
    3                +
                     |
                 XX  |
                    T|T A2_9_Emilie
                     |
    2           XXX  +
                     |
                XXX  |
        XXXXXXXXXXX  |
                     |S
    1         XXXXX  +
                    S|
                 XX  |
             XXXXXX  |
                     |  A2_10_Joseph  A2_2_Ana
    0      XXXXXXXX  +M
                     |  A2_5_Petra    A2_8_Leon
       XXXXXXXXXXXX  |
      XXXXXXXXXXXXX  |
                  X M|
   -1    XXXXXXXXXX  +
                     |S A2_7_Lena
          XXXXXXXXX  |  A2_3_Carlos
                     |
        XXXXXXXXXXX  |
   -2  XXXXXXXXXXXX  +
                     |
            XXXXXXX S|T
                     |
                 XX  |
   -3                +
            XXXXXXX  |
                 XX  |
                     |
                     |
   -4             X T+
                     |
                     |
                  X  |
                     |
   -5                +
                     |
                 XX  |
                     |
                     |
   -6                +
               <less>|<freq>
```

Visually checking the map, we see that the persons appear to be fairly well spread; however, the items are falling on the higher end of the continuum with most clustering closer to the middle.

That being said, we do notice some distinct levels, but with a large gap between the hardest item (A2_9, Emilie) and the rest of the items. Looking at the person and item separation and reliability statistics, we see that the person statistics are adequate with a separation index of 2.10 and a reliability estimate of 0.81, while the items statistics look really good with a separation index of 7.88 and a reliability estimate of 0.98. Finally, when comparing the ordering of the items in the variable map to the expected ordering in *Table 23*, we notice that there is one item that is not as expected—Joseph (A2_10)—which is falling quite a bit lower than theorized and creating the large gap by effectively eliminating the second-highest level.

Next, we look at the category characteristic curves (CCCs) shown in *Figure 25*, below, to determine whether the response options function as intended.

Figure 25

Comparative Math Anxiety (Mental-Cognitive) CCCs, R3

```
        CATEGORY PROBABILITIES: MODES - Andrich thresholds at intersections
P      -+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-
R  1.0 +                                                             +
O      |                                                             |
B      |                                                             |
A      |11                                                         55|
B   .8 +  11                                                     55  +
I      |    11                                                 55    |
L      |      1                                               5      |
I      |       1                   3333333                   5       |
T   .6 +        1                33       33                5        +
Y      |         11    222      3           3      44     55         |
    .5 +           1 22   222  3             3  444  444 5           +
O      |           2*        2*               **        *4           |
F   .4 +         22  1      33 2             4  3      5  44         +
       |        2     1    3    2           4    3    5     4        |
R      |      22       1  3      22       44      3  5       44      |
E      |     2          13         2     4         35          4     |
S   .2 +   22          3311         22 44         5533          44   +
P      |222           3    1         4*2         5    3           444|
O      |           333      11    444   222    55      333           |
N      |       3333           ****1       5****           3333       |
S   .0 +**********************55555*******11111**********************+
E      -+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-
       -5    -4    -3    -2    -1     0     1     2     3     4     5
        PERSON [MINUS] ITEM MEASURE
```

Again, the curves look very good. The response options are ordered as they should be and are all dominating in a good portion of the person minus item measure distribution. Then, we look at the item and person misfit tables (*Tables 27* and *28*, respectively). From *Table 27*, below, we see that item A2_9 (Emilie) has an INFIT MNSQ greater than 1.4 and an OUTFIT MNSQ greater than 1.2, while item A2_3 (Carlos) has both an INFIT and OUTFIT MNSQ greater than 1.2.

Table 27

Comparative Math Anxiety (Mental-Cognitive) Item Fit Statistics, R3

Item	Measure	INFIT		OUTFIT	
		MNSQ	*ZSTD*	*MNSQ*	*ZSTD*
A2_9	2.50	1.42	2.9	1.35	1.6
A2_3	-1.34	1.36	2.6	1.35	2.6
A2_7	-1.17	1.11	0.9	1.13	1.1
A2_2	0.16	0.82	-1.5	0.88	-0.9
A2_10	0.23	0.83	-1.4	0.81	-1.6
A2_5	-0.24	0.75	-2.2	0.73	-2.3
A2_8	-0.13	0.71	-2.6	0.68	-2.8

Overall, the misfit is not too high nor is it unexpected given what we have seen in previous rounds of data collection with the hardest and easiest items. From *Table 28*, on the following page, we see 30 out of 130 persons (approx. 23%) had an INFIT and/or OUTFIT MNSQ greater than or equal to 1.4, while an additional six students had INFIT and/or OUTFIT MNSQ values greater than 1.2 but less than 1.4.

Table 28

Comparative Math Anxiety (Mental-Cognitive) Person Fit Statistics, R3

Person	Measure	INFIT		OUTFIT	
		MNSQ	*ZSTD*	*MNSQ*	*ZSTD*
121	-0.31	5.91	4.4	5.74	4.3
60	1.96	4.48	4.2	5.05	4.6
92	1.64	4.26	4.1	4.44	4.2
27	-2.05	3.61	3.3	3.74	3.3
41	-3.49	1.72	1.3	3.59	2.4
103	1.96	3.52	3.4	3.41	3.3
47	1.33	3.19	3.1	3.20	3.1
64	-0.67	3.18	2.7	3.20	2.7
118	-0.67	2.87	2.4	2.81	2.4
18	1.33	2.37	2.2	2.36	2.2
110	-1.71	2.15	1.8	1.93	1.5
107	-2.39	1.64	1.3	2.04	1.6
15	0.37	1.94	1.6	1.91	1.5
160	-2.05	1.91	1.6	1.87	1.5
97	-0.67	1.86	1.4	1.88	1.4
150	1.64	1.78	1.5	1.82	1.6
1	2.63	1.54	1.1	1.81	1.5
90	0.03	1.73	1.3	1.63	1.1
2	-3.10	1.72	1.4	1.47	0.8
88	-5.35	1.21	0.5	1.65	0.9
113	-1.71	1.65	1.2	1.57	1.1
83	-1.02	1.64	1.2	1.61	1.1
111	-0.67	1.58	1.1	1.59	1.1
7	1.33	1.51	1.1	1.52	1.1
38	-1.02	1.50	1.0	1.44	0.9
93	1.33	1.46	1.0	1.47	1.0
20	0.37	1.45	0.9	1.41	0.8
76	-0.67	1.41	0.8	1.40	0.8
81	-2.39	0.99	0.1	1.41	0.9
9	-2.39	1.40	0.9	1.31	0.7
114	1.02	1.37	0.8	1.33	0.8
108	-0.31	1.34	0.7	1.32	0.7
57	0.03	1.32	0.7	1.31	0.7
28	-3.10	1.30	0.7	1.17	0.5
74	-1.37	1.15	0.5	1.25	0.6
115	-1.02	1.18	0.5	1.23	0.6

To further understand and support our previous interpretations of the item misfit, we look at the data in *Table 29*. Again, because several items were deleted, the items (and their item order locations) are as follows: A2_2 (location 1), A2_3 (location 2), and A2_5 (location 3), with items A2_7 through A2_10 in locations 4-7. As we have seen before, the misfit for both items A2_9 and A2_3 appear to be because of unexpected responses due to potential matching of items to response options rather than comparison of items to the response options. Specifically, misfit for item A2_9 (item location 6) occurred for the ten students shown when students responded 5 "much more anxious than" instead of an expected 3 "as anxious as" or when a student responded 2 "less anxious than" but was expected to respond 1 "much less anxious than," and misfit for item A2_3 (item location 2) occurred for the ten students shown when they responded 1 "much less anxious than" or 2 "less anxious than" instead of 3 "as anxious as" or 4 "more anxious than." Again, this is not unexpected and was not too impactful on the overall item misfit.

Table 29

Comparative Math Anxiety (Mental-Cognitive) Poorly Fitting Persons, R3

Person	Response Statistic	Item						
		1	*2*	*3*	*4*	*5*	*6*	*7*
121	Obs	4	1	3	1	4	3	4
	Exp	2.8	3.4	3.0	3.3	2.9	1.8	2.8
	Z		-3		-3			
60	Obs	4	2	5	3	4	5	4
	Exp	3.8	4.5	4.0	4.4	3.9	2.8	3.7
	Z		-3		-2		3	
92	Obs	3	4	4	2	3	5	5
	Exp	3.6	4.3	3.8	4.3	3.8	2.7	3.6
	Z				-3		3	2
27	Obs	3	1	2	1	4	2	2
	Exp	2.0	2.7	2.2	2.7	2.2	1.2	2.0
	Z		-2		-2	2		
41	Obs	2	1	2	1	1	2	2
	Exp	1.4	2.1	1.6	2.0	1.5	1.1	1.4
	Z						4	

Note. Obs = observed response, Exp = expected response, and Z = z-residual

Finally, we ran a principal component analysis on the Rasch residuals. Similar to the analysis of the emotional-attitudinal residuals, the determinant is non-zero at 0.025, the KMO is relatively low at 0.079, and Bartlett's Test of Sphericity is highly significant with X^2 (21) = 463.341, $p < .0005$. All of these statistics point to the fact that there is still some variability left to be explained in the item residuals. *Table B3*, in Appendix B, shows that there are four components with eigenvalue greater than 1; however, the scree plot (*Figure B3*) shows one clear component remaining. From *Table B4* with the component loadings, we see that items A2_3 and A2_7 (the two easiest items) have negative component loadings, while the other items have positive loadings. While not desirable, this is not unexpected as we already have evidence to suggest the extreme items, and in particular the easiest items, may be seen and responded to slightly differently than the other items. From *Figure B4*, the two component plot in rotated space, we again see that the residuals are randomly scattered for the most part with some items still correlated (e.g., z1, z5, and z7 or items A2_2, A2_8, and A2_10). This example contains three of the four items with moderate difficulty. So it appears items with similar difficulty may have other things in common, which explains the remaining common variability in the residuals.

Physical-somatic scale. The variable map for the physical-somatic scale from the final round of data collection can be found in *Figure 26*, on the following page.

Figure 26

Comparative Math Anxiety (Physical-Somatic) Variable Map, R3

```
MEASURE    PERSON - MAP - ITEM
               <more>|<rare>
    2                +  A3_3_Laila
                  X  |
                     |
                     |S
                 XX  |
    1            XX T+  A3_1_Ivan       A3_2_Gabrielle
                     |  A3_9_Noah
                XXX  |
               XXXX  |  A3_10_Erica
               XXXX  |
    0         XXXXX  +M
                    S|
           XXXXXXXX  |
          XXXXXXXXX  |
       XXXXXXXXXXXX  |  A3_7_Nora
   -1      XXXXXXXX  +
                     |
          XXXXXXXXX M|S
          XXXXXXXXX  |
           XXXXXXXX  |
   -2                +
         XXXXXXXXXX  |  A3_4_Jay        A3_8_Romita
      XXXXXXXXXXXXX  |
                    S|
        XXXXXXXXXXX  |T
   -3                +
                XXX  |
                     |
            XXXXXXX  |
                    T|
   -4                +
                     |
                  X  |
                     |
                     |
   -5                +
               <less>|<freq>
```

The persons and items appear to be fairly well spread; however, there is some mismatch between the persons and the items with the items trending towards the upper half of the person distribution. Looking at the separation indices and reliability estimates, we see that the person separation index is 1.67 and the person reliability estimate is 0.74 both of which are too low, while the item separation index is 10.07 and the item reliability estimate is 0.99 both are which are great. Lastly, when comparing the ordering of the items in the variable map to the expected ordering in *Table 23*, we see that most of the items are ordered as expected, with only one item—A3_2 (Gabrielle)—falling a little lower than we expected. That being said, when looking at the item difficulties (measure) in *Table 30*, below, we see the items are ordered as expected just not with as much distance between them as we would like.

Table 30

Comparative Math Anxiety (Physical-Somatic) Item Fit Statistics, R3

Item	Measure	INFIT		OUTFIT	
		MNSQ	*ZSTD*	*MNSQ*	*ZSTD*
A3_3	2.07	1.25	1.7	2.35	5.2
A3_8	-2.12	1.53	3.7	1.61	4.3
A3_4	-2.14	1.17	1.4	1.23	1.8
A3_1	0.95	0.98	-0.1	0.90	-0.8
A3_9	0.70	0.89	-0.9	0.87	-1.0
A3_10	0.35	0.88	-1.0	0.88	-1.0
A3_2	1.04	0.77	-2.0	0.74	-2.1
A3_7	-0.85	0.57	-4.3	0.57	-4.2

Next, we look at the category characteristic curves (CCCs) shown in *Figure 27*, on the following page, to determine whether the response options function as intended. Note that the response options are ordered and dominant in the appropriate portion of the distribution. However, it does appear that the fourth response option is being utilized more than expected while the second response option is being selected less than expected.

Figure 27

Comparative Math Anxiety (Physical-Somatic) CCCs, R3

```
        CATEGORY PROBABILITIES: MODES - Andrich thresholds at intersections
P      -+-----+-----+-----+-----+-----+-----+-----+-----+-----+-
R  1.0 +                                                       +
O      |                                                       |
B      |                                                       |
A      |                                                    555|
B   .8 +                                                   5   +
I      |11                                               55    |
L      |  1                                             5      |
I      |   1                                          55       |
T   .6 +    11                            44444      5         +
Y      |      1             3333333     44     44   5          |
    .5 +       1 222222    3       33 44         445           +
O      |       2*      2233          *            54           |
F   .4 +      2  1      322         4 3          5  44         +
       |    22    1    3   2       4   33       5     4        |
R      |   2       1  3     2    44      3    55       44      |
E      | 22         *3       22 4         3  5           4     |
S   .2 +2          3 1         *           3*             44   +
P      |         33   11     44 22        55 33             444|
O      |       33       11 44     22   555     33              |
N      |   3333       4444*111     5***22        33333         |
S   .0 +**************55555555*****111111**********************+
E      -+-----+-----+-----+-----+-----+-----+-----+-----+-----+-
       -4    -3    -2    -1     0     1     2     3     4     5
        PERSON [MINUS] ITEM MEASURE
```

Then, we look at *Tables 30* and *31* for the item and person misfit statistics, respectively. From *Table 30*, on the previous page, we see that the hardest item, A3_3, has an extremely high OUTFIT MNSQ (2.35) but the INFIT MNSQ is only 1.25, indicating most of the responses are consistent but with a significant amount of extreme responses. In addition, the two easiest items also have some misfit issues, where item A3_8 has an INFIT and OUTFIT MNSQ greater than 1.4 and item A3_4 has an OUTFIT MNSQ greater than 1.2. Again, some misfit is not unexpected with these items given the unique response options and the likelihood of that impacting the extreme ends of the scale more. That being said, the extremely high OUTFIT MNSQ for item A3_3 is very unexpected. From *Table 31*, on the following page, we see 21 of

129 students (approx. 16%) had INFIT and/or OUTFIT MNSQ values greater than 1.4, and an additional four students had INFIT and/or OUTFIT MNSQ values greater than 1.2 but less than 1.4.

Table 31

Comparative Math Anxiety (Physical-Somatic) Person Fit Statistics, R3

Person	Measure	INFIT		OUTFIT	
		MNSQ	*ZSTD*	*MNSQ*	*ZSTD*
41	-4.47	4.03	2.6	9.90	4.7
15	0.18	8.01	6.3	8.80	6.8
121	-0.33	6.68	5.7	7.29	6.1
18	0.70	5.65	4.9	6.07	5.3
108	-0.82	4.49	4.3	4.57	4.4
90	0.43	4.45	4.1	4.50	4.2
47	0.43	3.29	3.1	3.44	3.3
88	-2.81	3.00	2.6	3.19	2.5
64	-1.59	2.83	2.8	2.67	2.6
103	-1.33	2.46	2.4	2.43	2.4
107	-1.87	2.32	2.2	2.46	2.3
4	-2.81	2.35	2.0	1.91	1.4
63	1.14	2.32	2.0	2.13	1.9
74	-1.07	2.19	2.1	2.31	2.2
115	-1.07	1.99	1.8	2.05	1.9
137	0.70	2.04	1.8	1.94	1.7
61	-2.16	1.80	1.5	1.53	1.1
113	-2.81	1.61	1.1	1.77	1.2
22	0.18	1.57	1.2	1.66	1.3
87	-0.33	1.41	0.9	1.45	1.0
89	-2.47	1.07	0.3	1.43	0.9
114	-0.57	1.32	0.8	1.32	0.8
93	-0.82	1.29	0.7	1.24	0.6
110	-0.82	1.29	0.7	1.24	0.6
112	-2.47	1.26	0.6	1.12	0.4

To better understand the item misfit, particularly for item A3_3, we again look at the responses for the students with the greatest misfit in *Table 32*, on the following page.

Table 32

Comparative Math Anxiety (Physical-Somatic) Poorly Fitting Persons, R3

Person	Response Statistic	Item							
		1	*2*	*3*	*4*	*5*	*6*	*7*	*8*
41	Obs	1	1	3	1	M	1	1	1
	Exp	1.1	1.1	1	1.8		1.8	1.1	1.1
	Z			9					
15	Obs	4	3	4	1	3	1	5	4
	Exp	2.7	2.6	2.1	4.1	3.6	4.1	2.8	3.0
	Z			2	-4		-4	3	
121	Obs	3	4	4	1	3	1	3	4
	Exp	2.4	2.4	1.8	3.9	3.3	3.9	2.5	2.7
	Z		2	3	-4		-4		
18	Obs	3	4	5	2	3	2	4	4
	Exp	3.0	2.9	2.4	4.3	3.8	4.3	3.1	3.2
	Z			3	-3		-3		
108	Obs	4	4	3	2	2	2	1	3
	Exp	2.1	2.1	1.6	3.7	3.1	3.7	2.3	2.5
	Z	2	2	2	-2		-2		

Note. Obs = observed response, Exp = expected response, Z = z-residual, and M = missing data.

Because items A3_5 and A3_6 were deleted, items A3_7 through A3_10 can be found in item order locations 5-8. It appears that for item A3_3 (Laila gets nauseous to the point of feeling like she is going to be sick), there are quite a few students that are giving unexpectedly high responses, as if they were matching the hard item to the higher response options. In particular, student 41 has a very high z-residual (9) as they were expected to give a response of 1 "much less anxious than" given they responded 1 to all of the other items but instead they responded 3 "as anxious as" and therefore creating a very high person OUTFIT MSNQ (9.9) as well. This student may be the reason for the very high OUTFIT MNSQ for item A3_3, especially when combined with the more moderate, expected examples of student misfit on this item. For items A3_8 (location 6) and A3_4 (location 4), the misfit is back to what we have seen with the other two dimensions in this round of data collection and with all three dimensions in Round 1 where students responded 1 "much less anxious than" or 2 "less anxious than" rather than the expected response, most commonly, of 4 "more anxious than." This speaks to the matching phenomenon we are seeing for some students.

Lastly, a principal component analysis was run on the Rasch residuals. Similar to the previous residual analyses, the determinant is non-zero at 0.047, the KMO is relatively low at 0.197, and Bartlett's Test of Sphericity is highly significant with X^2 (28) = 380.766, $p < .0005$. All of these statistics point to the fact that there is still some variability left to be explained in the item residuals. *Table B5*, in Appendix B, shows that there are three components with eigenvalue greater than 1; however, the scree plot (*Figure B5*) shows one clear component remaining. From *Table B6* with the component loadings, we see that items A3_4, A3_7, and A3_8 (the three easiest items) have negative component loadings, while the other items have positive loadings. From *Figure B6*, the two component plot in rotated space, we again see that the residuals are randomly scattered for the most part with some items still correlated (e.g., z4 and z6 or items A3_4 and A3_8). As with the other dimensions, it appears items with similar difficulty may have other things in common, which explains the remaining common variability in the residuals.

Research Question 1: Applying and Adapting the Conceptual Framework

The first sub-research question is: How can the tripartite conceptual framework of mathematics anxiety Dr. Caroline Vuilleumier and I developed be employed and adapted in the development of a Rasch-based scale? Throughout the course of the instrument development process, the mathematics anxiety conceptual framework was used to guide the process of item development and revision. After each round, consideration was also given to whether or not the framework should be adapted based on evidence from the data collected. However, given inconsistencies in some of the items across administrations and the complexity in developing the scale, it was determined in each round that it would be inappropriate to make adaptations to the conceptual framework. After three rounds of data collection, that decision remains the same. While much more is known about the concept of mathematics anxiety and how to measure it along a continuum from low to high using Rasch measurement principles, there is still more that needs to be understood. There is still difficulty in describing some of the levels for each dimension, where items are not stabilizing or are behaving differently than expected. In addition to making additional improvements to the instrument to support revisions to the framework, it would also be appropriate at this point to engage with additional content experts from the fields of psychology and mathematics education in thinking about the tripartite conceptual framework of mathematics anxiety.

Research Question 2: Assessing the Measurement Invariance of the Scale

The second sub-research question is: Is the scale invariant across grade, gender, and other student demographics? This question was answered using the final round of data collection, where the only other student demographic collected besides grade and gender was self-reported English-language proficiency (ELP). To assess the measurement invariance of the scale across these three demographics, DIF analyses were conducted using Winsteps. Because of the multiple comparisons happening for a single analysis, a Bonferroni-corrected alpha was used to determine significance. For the emotional-attitudinal and physical-somatic scales, which have eight items each, the Bonferroni-corrected alpha is 0.00625 (0.05/8), and for the mental-cognitive scale, which has seven items, the Bonferroni-corrected alpha is 0.00714 (0.05/7). Resulting output tables from the DIF analyses can be found in Appendix C.

Emotional-Attitudinal Scale

The DIF results for grade, gender, and ELP and the emotional-attitudinal scale can be found in *Tables C1, C2,* and *C3*, respectively. Using a Bonferroni-corrected alpha of 0.00625, no significant differences were found for grade and gender. However, three significant differences were found for ELP (see *Figure 28*, below, for the ELP DIF plot).

Figure 28

Comparative Math Anxiety (Emotional-Attitudinal) DIF Plot, English-Language Proficient

Specifically, the Rasch-Welch test of significance flagged item A1_5 (Ruby is terrified) with a *p-value* of 0.0051 and item A1_11 (Chloe feels unfazed) with a *p-value* of 0.0017, while the Mantel-Haenszel flagged item A1_10 (Andrea is distressed) with a *p-value* of 0.0013. For items A1_5 and A1_10, students who identified as English-language proficient found the items much harder (had a higher DIF measure) than students who did not identify as English-language proficient. In contrast, students who identified as English-language proficient found item A1_11 much easier (had a lower DIF measure) than students who did not identify as English-language proficient. Items A1_5 and A1_10 were theorized to be two of the hardest items, while A1_11 was theorized to be one of the two easiest items. Therefore, it appears students who identified as English-language proficient had an easier time interpreting and understanding the intention of these three items than students who did not identify as English-language proficient, and these items may be inappropriate for use with students who do not identify as English-language proficient.

Mental-Cognitive Scale

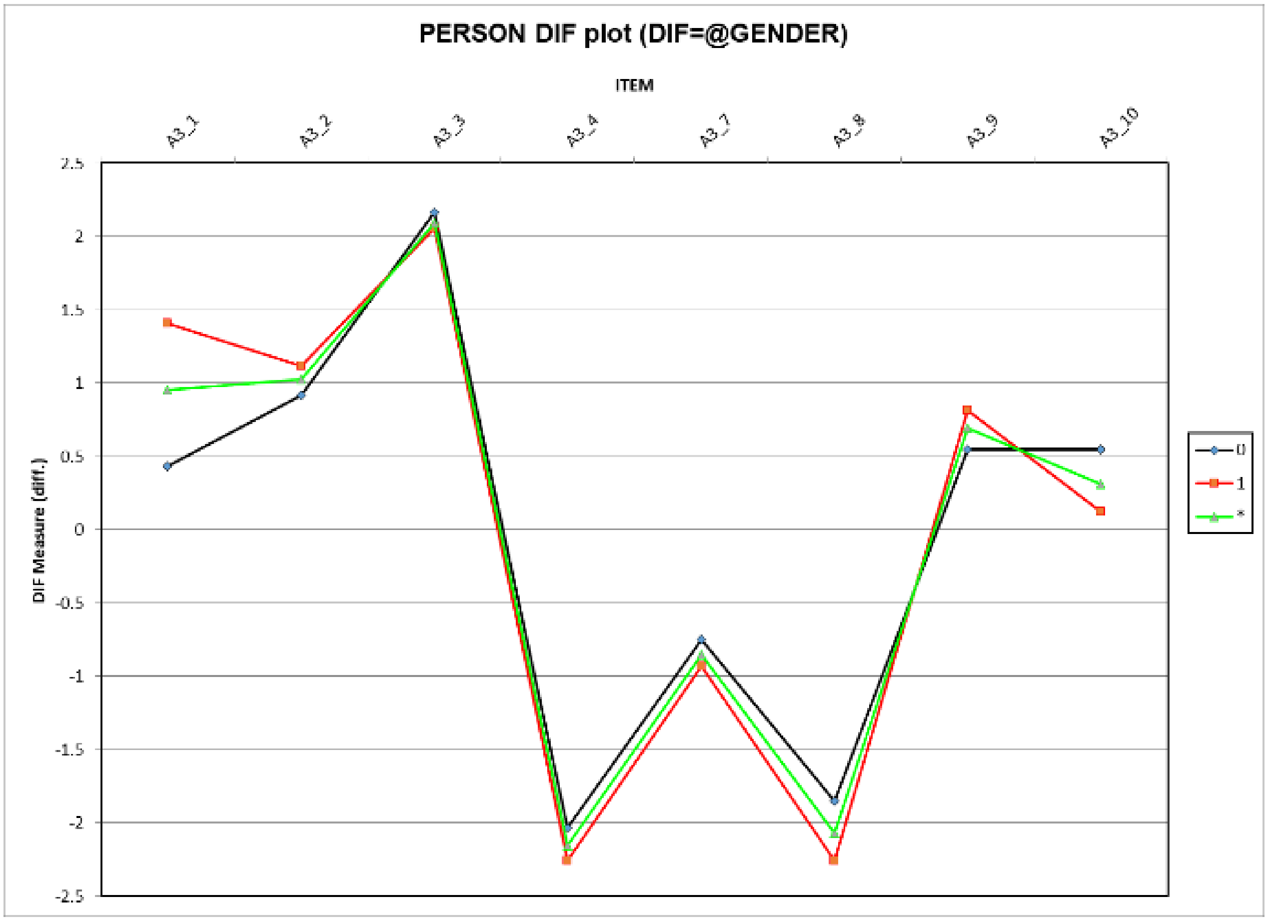
The DIF results for grade, gender, and ELP and the mental-cognitive scale can be found in *Tables C4, C5,* and *C6*, respectively. Using a Bonferroni-corrected alpha of 0.00714, no significant differences were found for any of the student demographics.

Physical-Somatic Scale

The DIF results for grade, gender, and ELP and the physical-somatic scale can be found in *Tables C7, C8,* and *C9*, respectively. Using a Bonferroni-corrected alpha of 0.00625, no significant differences were found for grade-level. However, differences were found for gender and ELP (see *Figure 29*, on the following page, and *Figure 30*, on page 148, for the gender and ELP DIF plots, respectively).

Figure 29

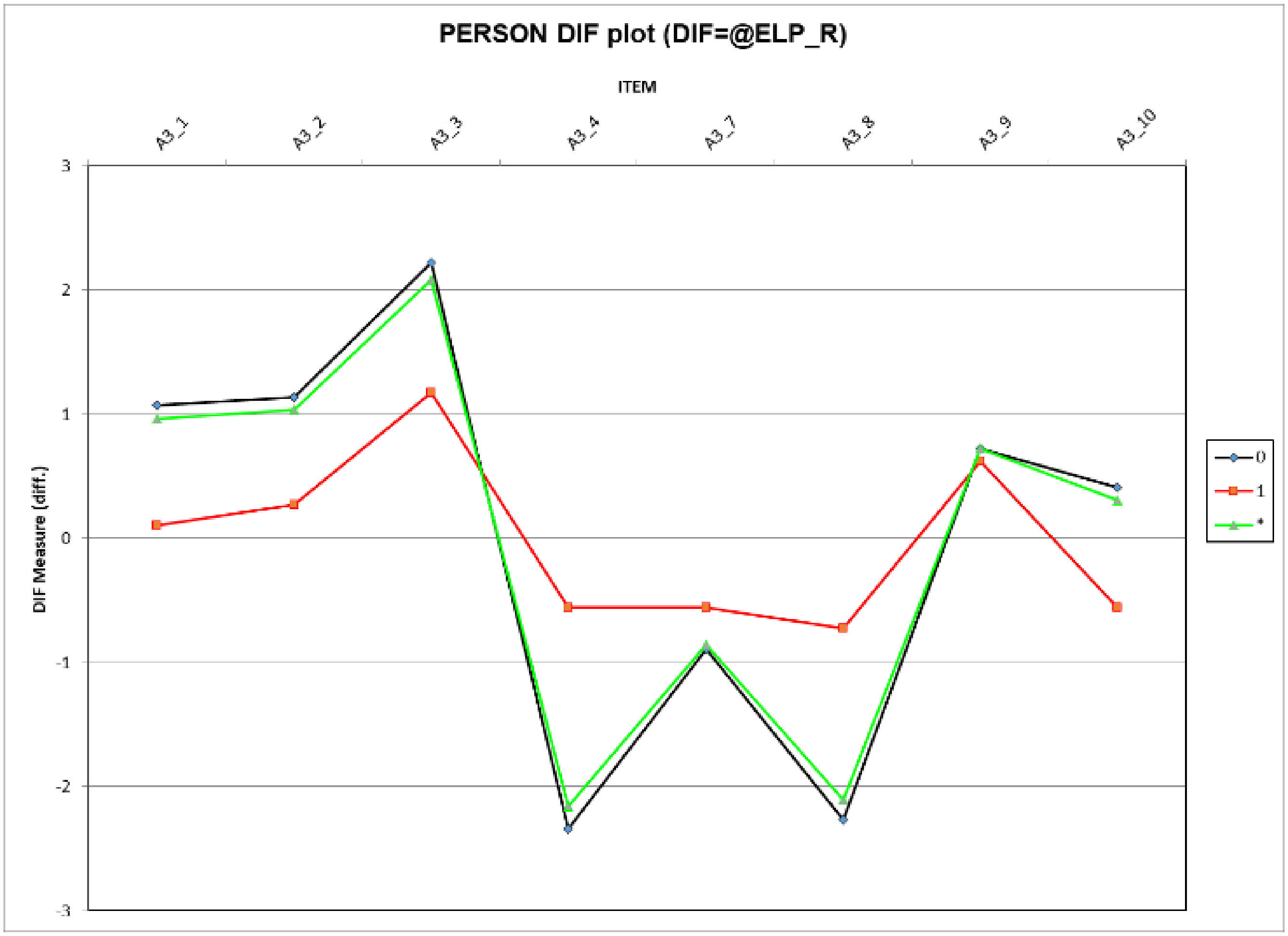
Comparative Math Anxiety (Physical-Somatic) DIF Plot, Gender



For gender, both the Rasch-Welch and Mantel-Haenszel tests of significance flagged item A3_1 (Ivan feels physically uncomfortable) with *p-values* of 0.0006 and 0.0003, respectively. We can see that for this item, male students found the item harder (had a higher DIF measure) than female students, indicating male students were more likely to select *much more anxious* or *more anxious* in comparison to female students, and female students were more likely to select *less anxious* or *much less anxious* in comparison to male students. It is unclear from data collected for this study as to why this difference may exist, but at this point, it may be appropriate to exclude this item from analysis. This would not be the greatest loss to construct measurement given the item's relative redundancy in location with items A3_2 and A3_9.

Figure 30

Comparative Math Anxiety (Physical-Somatic) DIF Plot, English-Language Proficient



For ELP, the Rasch-Welch test flagged two items—item A3_4 (Romita feels completely at ease) and item A3_8 (Jay is completely calm) with *p-values* of 0.0011 and 0.0033, respectively. For both items, students who identified as English-language proficient found the items much easier (had a lower DIF measure) than students who did not identify as English-language proficient. As with item A1_11 in the emotional-attitudinal subscale, these two items were theorized to be the easiest two items in the physical-somatic subscale and students who do not identify as English-language proficient seem to have had a harder time understanding the items and how to respond to them appropriately based on the directions. More work is needed to understand what changes may be needed to support students who do not identify as English-language proficient in appropriately responding to these items.

Research Question 3: Construct and Criterion Validity Evidence

The third sub-research question is: Does the scale have an acceptable degree of construct and criterion validity? To provide evidence for the construct validity of the scale, we compared the variable maps to the *a prior* conceptualization of mathematics anxiety. This was done throughout the scale development process and has already been described in this chapter—most recently in the Round 3 Results section. As a reminder, the emotional-attitudinal items ordered mostly as expected with the exception of items A1_3 and A1_7, which were hypothesized to both be in the second-lowest level but were harder than expected; item A2_10 from the mental-cognitive scale fell quite a bit lower than theorized, creating a large gap and effectively eliminating the second-highest level of the construct; and item A3_2 fell a little lower than expected in the physical-somatic scale. So while there is some evidence to support the construct validity of the three dimensions of the Comparative Mathematics Anxiety scale, each subscale could use more work to better represent the dimensions of mathematics anxiety. While the conceptual framework and scale items do not completely align, guidelines for use of the current version of the instrument should still be provided at this point (Step 7).

Construct Validity Evidence: Emotional-Attitudinal Anxiety

Figure 31, on the following page, presents the variable map for the final version of the emotional-attitudinal dimension of the Comparative Mathematics Anxiety Scale (CMAS-EA).

Figure 31

Structure of the Emotional-Attitudinal Dimension of the CMAS

```
MEASURE   PERSON - MAP - ITEM
(logits)
              <more>|<rare>
    3               +
                    |
                    |
                 X  |
                    |T
    2           XX T+  A1_5_Ruby                           5) terrified
               XXX  |
                 X  |
                    |
              XXXX  |
    1         XXXX  +S
              XXXX  |  A1_10_Andrea                        10) distressed
             XXXXX S|
                    |  A1_13_Larissa                       13) emotionally worked up
        XXXXXXXXXX  |  A1_3_Viktoria                       3) a bit apprehensive
    0       XXXXXX  +M
          XXXXXXXX  |  A1_12_Gabriel  A1_7_David           12) sense of worry
             XXXXX  |                                      7) a little intimidated
                 X  |
      XXXXXXXXXXXX M|
   -1       XXXXXX  +S
          XXXXXXXX  |
                 X  |  A1_4_Daniel                         4) in element
            XXXXXX  |  A1_11_Chloe                         11) unfazed
          XXXXXXXX  |
   -2      XXXXXXX  +
                   S|T
         XXXXXXXXX  |
          XXXXXXXX  |
                    |
   -3         XXXX  +
                 X  |
                 X  |
                   T|
                    |
   -4               +
                    |
                 X  |
                    |
                    |
   -5               +
              <less>|<freq>
```

As stated before, if our measurement purposes are met, we would ideally like to see a “ladder-like” progression of items, such that items are relatively uniformly spaced out along the continuum from lower to higher levels of anxiety, and we would like the persons to be fairly distributed and match across to the items. Most important, however, we want the empirical results for the item difficulty estimates to make sense. That is, do we have evidence of a construct valid anxiety scale that defines a meaningful progression of items along a continuum of easier-to-harder to be “much more anxious than X” item?

Starting at the bottom of *Figure 31* for the emotional-attitudinal anxiety variable it is easiest to be “much more anxious than X” on items A1_11 (unfazed) and A1_4 (in element). These two items are followed by moderately harder (more negative) items A1_12 (sense of worry) and A1_7 (a little intimidated). Slightly above the middle of the distribution of items is item A1_3 (a bit apprehensive), which is quickly followed by slightly harder items A1_13 (emotionally worked up) and A1_10 (distressed). Finally, at the top of the variable map, item A1_5 (terrified) defines the highest negative item on the emotional-attitudinal anxiety variable because it is the hardest to be “much more anxious than.” It is clear that increasing levels of emotional-attitudinal mathematics anxiety are being captured by the items as they rise from the bottom to the top of the scale. Consistent with our Rasch scale development expectations, more extreme indicators of emotional and attitudinal mathematics anxiety are being exhibited as we proceed up the emotional-attitudinal dimension of the Comparative Mathematics Anxiety Scale.

While *Figure 31* shows the mean location for each item based on its total score across all students, it does not reveal the location of different response categories, or the level of response, expected of a student to any one item at a given location. One way to reveal expected responses along the continuum is presented in *Figure 32*, on the following page. Using the score equivalence table provided in Winsteps (Table 20.1), we can find how any one student's raw score on the scale translates into an anxiety level location on this map.

Figure 32

Translation to Raw Scores for the Emotional-Attitudinal Dimension of the CMAS

```
MEASURE   PERSON - MAP – ITEM
 (raw)
   40          <more>|<rare>                                    Average score of 5
   35               +
                    |
   34               |
   33             X |
                    |T                                          Average score of 4
   32            XX T+  A1_5_Ruby
   31           XXX |
   30             X |
                    |
   29          XXXX |
   28          XXXX +S
   27          XXXX |  A1_10_Andrea
   26         XXXXX S|
                    |  A1_13_Larissa
   25   XXXXXXXXXXX |  A1_3_Viktoria
   24        XXXXXX +M                                          Average score of 3
   23      XXXXXXXX |  A1_12_Gabriel  A1_7_David
   22         XXXXX |
                  X |
   21 XXXXXXXXXXXXX M|
   20        XXXXXX +S
   19      XXXXXXXX |
                  X |  A1_4_Daniel
   18        XXXXXX |  A1_11_Chloe
   17      XXXXXXXX |
   16       XXXXXXX +                                           Average score of 2
                   S|T
   15     XXXXXXXXX |
   14      XXXXXXXX |
                    |
   13          XXXX +
                  X |
   12             X |
                   T|
   11               |
                    +
                    |
   10             X |
                    |
    9               |
    8               +                                           Average score of 1
               <less>|<freq>
```

For example, if a person had a raw score of 28, their "measure" (or logit score) would be 1.02, and they would be represented as one of the "X" marks adjacent to the "1" along the left side of the variable map. Specifically, the horizontal lines in *Figure 32* mark the sections along the mathematics anxiety variable that correspond to average response scores of 1 through 5. No one had a total score that corresponded to an average of 1 on the items. The largest concentration of students occurs between the averages of 2 and 3. There is a small group of students who had average scores between 1 and 2 and another relatively small group of students who had average scores between 3 and 4, with one student averaging between 4 and 5.

While *Figure 32* is useful for understanding how the logit scores translate to raw scores and how these scores relate to use of the response options, it does not provide the most useful interpretation of the scale from a clinical perspective. That is, what does an average score of 3 or 4 actually mean? And, is there a more useful way to interpret these scores to understand a student's level of emotional-attitudinal anxiety? *Figure 33*, on the following page, provides a more explicit, clear interpretation of what the different scores mean related to levels of emotional-attitudinal anxiety.

Figure 33

Interpretation of the Emotional-Attitudinal Dimension of the CMAS

```
MEASURE   PERSON - MAP – ITEM
 (raw)
   40          <more>|<rare>                                      Extremely high anx.
   35                +
                     |
   34                |
   33             X  |
                     |T
   32            XX T+  A1_5_Ruby                           5) terrified
   31           XXX  |
   30             X  |                                            Very high anxiety
                     |
   29          XXXX  |
   28          XXXX  +S                                     10) distressed
   27          XXXX  |  A1_10_Andrea                        13) emotionally worked up
   26         XXXXX S|                                      3) a bit apprehensive
                     |  A1_13_Larissa
   25    XXXXXXXXXX  |  A1_3_Viktoria                             High anxiety
   24        XXXXXX  +M
   23      XXXXXXXX  |  A1_12_Gabriel  A1_7_David           12) sense of worry
   22         XXXXX  |                                      7) a little intimidated
                  X  |
   21 XXXXXXXXXXXX M|
   20        XXXXXX  +S                                           Moderate anxiety
   19      XXXXXXXX  |
                  X  |  A1_4_Daniel                         4) in element
   18        XXXXXX  |  A1_11_Chloe                         11) unfazed
   17      XXXXXXXX  |
   16       XXXXXXX  +                                            Low anxiety
                    S|T
   15     XXXXXXXXX  |
   14      XXXXXXXX  |
                     |
   13          XXXX  +
                  X  |
   12             X  |
                    T|
   11                |
                     +
                     |
   10             X  |
                     |
    9                |                                            Very low anxiety
    8                +                                            No anxiety
               <less>|<freq>
```

For ease of practical application, *Table 33*, below, reproduces the information visually shown in *Figure 33*, including score ranges and interpretations that may be utilized by any person who responds to the CMAS-EA.

Table 33

Emotional-Attitudinal Scale Score Translation Summary

Score	Anxiety Level	Description of Score	Item
40	Extremely high anxiety	You are "much more anxious than" every item presented.	
30-39	Very high anxiety	On average, you are "about as anxious as" or "more anxious than" Item A1_5 and "much more anxious than" all items below this section.	A1_5) Terrified
25-29	High anxiety	On average, you are "about as anxious as" or "more anxious than" Items A1_10, A1_13, and A1_3, "more anxious than" or "much more anxious than" the items below this section, but "less anxious than" Item A1_5.	A1_10) Distressed A1_13) Emotionally worked up A1_3) A bit apprehensive
20-24	Moderate anxiety	On average, you are "about as anxious as" Items A1_12 and A1_7, "more anxious than" the items below this section, but "less anxious than" the items above this section.	A1_12) Sense of worry A1_7) A little intimidated
16-19	Low anxiety	On average, you are "about as anxious as" Items A1_4 and A1_11 and "much less anxious than" or "less anxious than" the items above this section.	A1_4) In element A1_11) Unfazed
9-15	Very low anxiety	On average, you are "less anxious than" or "much less anxious than" every item presented.	
8	No anxiety	You are "much less anxious than" every item presented.	

In an intervention situation designed to reduce students' level of anxiety, these qualitative interpretations based on a person's estimated location on the variable can be useful for charting changes in levels of emotional-attitudinal mathematics anxiety over time. If desired, a user of these scales could utilize these score ranges to develop descriptive terms that represented qualitatively different levels on the CMAS-EA (Wilson, 2005). Similar interpretations will now be provided for the mental-cognitive and physical-somatic dimensions of the Comparative Mathematics Anxiety Scale.

Construct Validity Evidence: Mental-Cognitive Anxiety

The variable map for the final version of the mental-cognitive dimension of the Comparative Mathematics Anxiety Scale (CMAS-MC) can be found in *Figure 34*, on the following page. While we want to see a "ladder-like" progression of items and persons spread out along the continuum, the most important thing is that the empirical results for the item difficulty estimates make sense. Starting at the bottom of *Figure 34* for the mental-cognitive anxiety variable, it is easiest to be "much more anxious than X" on items A2_3 (think clearly) and A2_7 (focused). These two items are followed slightly below the middle of the distribution of items by items A2_8 (thoughts are scattered) and A2_5 (confused), while slightly above the middle of the distribution of items are item A2_2 (can't recall what learned) and A2_10 (mind keeps blanking out). There is a large gap, and then finally, at the top of the variable map, item A2_9 (think about nothing else except fear of math) defines the highest negative item on the mental-cognitive anxiety variable because it is the hardest to be "much more anxious than." It is clear that increasing levels of mental-cognitive mathematics anxiety are being captured by the items as they rise from the bottom to the top of the scale. Consistent with our Rasch scale development expectations, more extreme indicators of mental and cognitive mathematics anxiety are being exhibited as we proceed up the mental-cognitive dimension of the Comparative Mathematics Anxiety Scale.

Figure 34

Structure of the Mental-Cognitive Dimension of the CMAS

```
MEASURE    PERSON - MAP – ITEM
(logits)
               <more>|<rare>
    3                +
                     |
                 XX  |
                    T|T A2_9_Emilie                      9) think about... fear
                     |
    2           XXX  +
                     |
                XXX  |
        XXXXXXXXXXX  |
                     |S
    1         XXXXX  +
                    S|
                 XX  |
             XXXXXX  |
                     |  A2_10_Joseph  A2_2_Ana           10) mind... blanking out
    0      XXXXXXXX  +M                                  2) can’t recall... learned
                     |  A2_5_Petra    A2_8_Leon          5) confused
       XXXXXXXXXXXX  |                                   8) thoughts are scattered
      XXXXXXXXXXXXX  |
                  X M|
   -1    XXXXXXXXXX  +
                     |S A2_7_Lena                        7) focused
          XXXXXXXXX  |  A2_3_Carlos                      3) think clearly
                     |
        XXXXXXXXXXX  |
   -2  XXXXXXXXXXXX  +
                     |
            XXXXXXX S|T
                     |
                 XX  |
   -3                +
            XXXXXXX  |
                 XX  |
                     |
                     |
   -4             X T+
                     |
                     |
                  X  |
                     |
   -5                +
                     |
                 XX  |
                     |
                     |
   -6                +
               <less>|<freq>
```

Again, *Figure 34* does not reveal the location of different response categories, or the level of response, expected of a student to any one item at a given location. One way to reveal expected responses along the continuum is presented in *Figure 35*, on the following page, again using information from Winstep's score equivalence table (20.1). Here, if a person had a raw score of 24, their "measure" (or logit score) would be 1.02, and they would be represented as one of the "X" marks adjacent to the "1" along the left side of the variable map. The horizontal lines in *Figure 35* mark the sections along the mathematics anxiety variable that correspond to average response scores of 1 through 5. Very similar to the emotional-attitudinal mathematics anxiety variable, no one had a total score that corresponded to an average of 1 on the items; the largest concentration of students occurs between the averages of 2 and 3; there is a small group of students who had average scores between 1 and 2 and another relatively small group of students who had average scores between 3 and 4; and there are two students (instead of one student) averaging between 4 and 5.

Figure 35

Translation to Raw Scores for the Mental-Cognitive Dimension of the CMAS

```
MEASURE    PERSON - MAP – ITEM
 (raw)
   35           <more>|<rare>                                           Average score of 5
   30                 +
                      |
   29              XX |
   28                T|T A2_9_Emilie                                    Average score of 4
                      |
   27             XXX +
                      |
                  XXX |
        XXXXXXXXXXX   |
                      |S
   24           XXXXX +
                     S|
   23              XX |
   22          XXXXXX |
                      |  A2_10_Joseph  A2_2_Ana
   21        XXXXXXXX +M                                                Average score of 3
                      |  A2_5_Petra    A2_8_Leon
   20  XXXXXXXXXXXX   |
   19 XXXXXXXXXXXXX   |
                   X M|
   18     XXXXXXXXXX  +
                      |S A2_7_Lena
   17      XXXXXXXXX  |  A2_3_Carlos
                      |
   16   XXXXXXXXXXX   |
   15  XXXXXXXXXXXX   +
                      |
   14         XXXXXXX S|T                                               Average score of 2
                      |
   13              XX |
                      +
   12         XXXXXXX |
   11              XX |
                      |
                      |
   10               X T+
                      |
                      |
    9               X |
                      |
                      +
                      |
    8              XX |
                      |
                      |
    7                 +                                                 Average score of 1
                <less>|<freq>
```

While *Figure 35* is useful for understanding how the logit scores translate to raw scores and how these scores relate to use of the response options, it does not provide the most useful interpretation of the scale from a clinical perspective. *Figure 36*, below, provides a more explicit interpretation of what the different scores mean related to levels of mental-cognitive anxiety.

Figure 36

Interpretation of the Mental-Cognitive Dimension of the CMAS

```
MEASURE    PERSON - MAP – ITEM
 (raw)
   35          <more>|<rare>                                               Extremely high anx.
   30                +
                     |
   29             XX  |                                   9) think about... fear
   28               T|T A2 9 Emilie                                      Very high anxiety
                     |
   27            XXX  +
                     |
                 XXX  |
        XXXXXXXXXXX  |
                     |S
   24          XXXXX  +
                    S|
   23             XX  |                                   10) mind... blanking out
   22         XXXXXX  |                                   2) can’t recall... learned
                     |  A2 10 Joseph  A2 2 Ana                   High anxiety
   21       XXXXXXXX  +M                                  5) confused
                     |  A2_5_Petra    A2_8_Leon           8) thoughts are scattered
   20  XXXXXXXXXXXX  |
   19 XXXXXXXXXXXXX  |                                                   Moderate anxiety
                  X M|
   18    XXXXXXXXXX  +
                     |S A2_7_Lena                         7) focused
   17     XXXXXXXXX  |  A2_3_Carlos                       3) think clearly
                     |
   16   XXXXXXXXXXX  |
   15  XXXXXXXXXXXX  +
                     |                                                   Low anxiety
   14        XXXXXXX S|T
                     |
   13             XX  |
                     +
   12        XXXXXXX  |
   11             XX  |
                     |
                     |
   10              X T+
                     |
                     |
    9              X  |
                     |
                     +
                     |
    8             XX  |                                                   Very low anxiety
                     |
                     |
    7                +                                                   No anxiety
               <less>|<freq>
```

For ease of practical application, *Table 34*, below, reproduces the information visually shown in *Figure 36*, including score ranges and interpretations that may be utilized by any person who responds to the CMAS-MC.

Table 34

Mental-Cognitive Scale Score Translation Summary

Score	Anxiety Level	Description of Score	Item
35	Extremely high anxiety	You are "much more anxious than" every item presented.	
28-34	Very high anxiety	On average, you are "about as anxious as" or "more anxious than" Item A2_9 and "much more anxious than" all items below this section.	A2_9) Think about nothing else except fear of math
22-27	High anxiety	On average, you are "about as anxious as" or "more anxious than" Items A2_10 and A2_2, "more anxious than" or "much more anxious than" the items below this section, but "less anxious than" Item A2_9.	A2_10) Mind keeps blanking out A2_2) Can't recall what learned
19-21	Moderate anxiety	On average, you are "about as anxious as" Items A2_5 and A2_8, "more anxious than" the items below this section, but "less anxious than" the items above this section.	A2_5) Confused A2_8) Thoughts are scattered
15-18	Low anxiety	On average, you are "about as anxious as" Items A2_7 and A2_3 and "much less anxious than" or "less anxious than" the items above this section.	A2_7) Focused A2_3) Think clearly
8-14	Very low anxiety	On average, you are "less anxious than" or "much less anxious than" every item presented.	
7	No anxiety	You are "much less anxious than" every item presented.	

Construct Validity Evidence: Physical-Somatic Anxiety

The variable map for the final version of the physical-somatic dimension of the Comparative Mathematics Anxiety Scale (CMAS-PS) can be found in *Figure 37.* Again, while we want to see a "ladder-like" progression of items and persons spread out along the continuum, the most important thing is that the empirical results for the item difficulty estimates make sense.

Figure 37

Structure of the Physical-Somatic Dimension of the CMAS

```
MEASURE    PERSON - MAP – ITEM
(logits)
               <more>|<rare>
    3                +
                     |
                     |
                     |
                     |
    2                +  A3_3_Laila                            3) nauseous
                  X  |
                     |
                     |S
                 XX  |                                        1) uncomfortable
    1            XX T+  A3_1_Ivan       A3_2_Gabrielle        2) heart is racing
                     |  A3_9_Noah                             9) palms start to sweat
                XXX  |
               XXXX  |  A3_10_Erica                           10) headache
               XXXX  |
    0         XXXXX  +M
                    S|
           XXXXXXXX  |
          XXXXXXXXX  |
       XXXXXXXXXXXX  |  A3_7_Nora                             7) a little tense
   -1      XXXXXXXX  +
                     |
          XXXXXXXXX M|S
          XXXXXXXXX  |
           XXXXXXXX  |
   -2                +
         XXXXXXXXXX  |  A3_4_Jay        A3_8_Romita           4) calm
      XXXXXXXXXXXXX  |                                        8) at ease
                    S|
        XXXXXXXXXXX  |T
   -3                +
                XXX  |
                     |
            XXXXXXX  |
                    T|
   -4                +
                     |
                  X  |
                     |
                     |
   -5                +
               <less>|<freq>
```

Starting at the bottom of *Figure 37* for the physical-somatic anxiety variable, it is easiest to be "much more anxious than X" on items A3_8 (at ease) and A3_4 (calm). These two items are followed by A3_7 (a little tense). Slightly above the middle of the distribution of items is item A3_10 (headache). Slightly above that are items A3_9 (palms start to sweat), A3_2 (heart is racing), and item A3_1 (physically uncomfortable). Finally, at the top of the variable map, item A3_3 (nauseous to the point of feeling like going to be sick) defines the highest negative item on the physical-somatic anxiety variable because it is the hardest to be "much more anxious than." Clearly, increasing levels of physical-somatic mathematics anxiety are being captured by the items as they rise from the bottom to the top of the scale. Consistent with our Rasch scale development expectations, more extreme indicators of physical and somatic mathematics anxiety are being exhibited as we proceed up the physical-somatic dimension of the Comparative Mathematics Anxiety Scale.

Again, *Figure 37* does not reveal the location of different response categories, or the level of response, expected of a student to any one item at a given location. *Figure 38*, on the following page, presents one way to reveal expected responses along the continuum, using information from Winstep's score equivalence table (20.1). Here, if a person had a raw score of 28, their "measure" (or logit score) would be 0.97, and they would be represented as one of the "X" marks adjacent to the "1" along the left side of the variable map. The horizontal lines in *Figure 38* mark the sections along the mathematics anxiety variable that correspond to average response scores of 1 through 5. Very similar to the other mathematics anxiety variables, no one had a total score that corresponded to an average of 1 on the items; the largest concentration of students occurs between the averages of 2 and 3; there is a small group of students who had average scores between 1 and 2 and another relatively small group of students who had average scores between 3 and 4; and there are zero students averaging between 4 and 5.

Figure 38

Translation to Raw Scores for the Physical-Somatic Dimension of the CMAS

```
MEASURE    PERSON - MAP – ITEM
 (raw)
   40          <more>|<rare>                                     Average score of 5
                     +
   34                |
   33                |
                     |
   32                |                                           Average score of 4
                     +  A3_3_Laila
   31             X  |
                     |
   30                |S
   29            XX  |
   28            XX T+  A3_1_Ivan       A3_2_Gabrielle
                     |  A3_9_Noah
   27           XXX  |
   26          XXXX  |  A3_10_Erica
   25          XXXX  |
   24         XXXXX  +M                                          Average score of 3
                    S|
   23      XXXXXXXX  |
   22     XXXXXXXXX  |
   21  XXXXXXXXXXXX  |  A3_7_Nora
   20      XXXXXXXX  +
                     |
   19     XXXXXXXXX M|S
   18     XXXXXXXXX  |
   17      XXXXXXXX  |
                     +
   16    XXXXXXXXXX  |  A3_4_Jay        A3_8_Romita              Average score of 2
   15 XXXXXXXXXXXXX  |
                    S|
   14   XXXXXXXXXXX  |T
                     +
   13           XXX  |
                     |
   12       XXXXXXX  |
                    T|
   11                +
                     |
   10             X  |
                     |
    9                |
    8                +                                           Average score of 1
               <less>|<freq>
```

Again, *Figure 38* does not provide the most useful, clinical interpretation of the scale. *Figure 39*, below, provides a more explicit interpretation of what the different scores mean related to levels of physical-somatic anxiety.

Figure 39

Interpretation of the Physical-Somatic Dimension of the CMAS

```
MEASURE    PERSON - MAP – ITEM
 (raw)
   40          <more>|<rare>                                          Extremely high anx.
                     +
   34                |
   33                |
                     |
   32                |
                     +  A3_3_Laila                           3) nauseous
   31             X  |
                     |
   30                |S                                             Very high anxiety
   29            XX  |                                       1) uncomfortable
   28            XX T+  A3_1_Ivan        A3_2_Gabrielle      2) heart is racing
                     |  A3_9_Noah                            9) palms start to sweat
   27           XXX  |                                              High anxiety
   26          XXXX  |  A3_10_Erica                          10) headache
   25          XXXX  |
   24         XXXXX  +M                                             Moderate-high anx.
                    S|
   23      XXXXXXXX  |
   22     XXXXXXXXX  |
   21  XXXXXXXXXXXX  |  A3_7_Nora                            7) a little tense
   20      XXXXXXXX  +                                              Moderate anxiety
                     |
   19     XXXXXXXXX M|S
   18     XXXXXXXXX  |
   17      XXXXXXXX  |
                     +                                       4) calm
   16    XXXXXXXXXX  |  A3_4_Jay         A3_8_Romita         8) at ease
   15 XXXXXXXXXXXXX  |                                              Low anxiety
                    S|
   14   XXXXXXXXXXX  |T
                     +
   13           XXX  |
                     |
   12       XXXXXXX  |
                    T|
   11                +
                     |
   10             X  |
                     |
    9                |                                              Very low anxiety
    8                +                                              No anxiety
               <less>|<freq>
```

For ease of practical application, *Table 35*, below, reproduces the information visually shown in *Figure 39*, including score ranges and interpretations that may be utilized by any person who responds to the CMAS-PS.

Table 35

Physical-Somatic Scale Score Translation Summary

Score	Anxiety Level	Description of Score	Item
40	Extremely high anxiety	You are "much more anxious than" every item presented.	
30-39	Very high anxiety	On average, you are "about as anxious as" or "more anxious than" Item A3_3 and "much more anxious than" all items below this section.	A3_3) Nauseous to the point of feeling like going to be sick
27-29	High anxiety	On average, you are "about as anxious as" Items A3_1, A3_2, and A3_9, "more anxious than" or "much more anxious than" the items below this section, but "less anxious than" Item A3_3.	A3_1) Physically uncomfortable A3_2) Heart is racing A3_9) Palms start to sweat
24-26	Moderately-high anxiety	On average, you are "about as anxious as" Item 3_10, "more anxious than" or "much more anxious than" the items below this section, but "less anxious than" or "much less anxious than" the items above this section.	A3_10) Headache
20-23	Moderate anxiety	On average, you are "about as anxious as" Item A3_7, "more anxious than" the items below this section, but "less anxious than" the items above this section.	A3_7) A little tense
15-19	Low anxiety	On average, you are "about as anxious as" Items A3_4 and A3_8 and "much less anxious than" or "less anxious than" the items above this section.	A3_4) Calm A3_8) At ease
9-14	Very low anxiety	On average, you are "less anxious than" or "much less anxious than" every item presented.	
8	No anxiety	You are "much less anxious than" every item presented.	

Criterion Validity Evidence

To provide evidence for the criterion validity of the scale, we want to look at the correlations between students' scores on the three dimensions of the Comparative Mathematics Anxiety Scale and their scores on the other measures collected on the third and final survey. First, we want to see strong, positive correlations between the three dimensions. Then, as evidence of the convergent validity of the three subscales, there should be a moderately-strong positive correlation between student scores on the Comparative Mathematics Anxiety subscales and their scores on the mAMAS (Carey et al., 2017), as well as a positive, but comparatively weaker, correlation with their fixed mindset scores. Finally, evidence of divergent validity exists if there are negative correlations between the mathematics anxiety subscales and the following mathematics constructs: confidence, motivation, perseverance, and growth mindset. Note that the Comparative Mathematics Anxiety Scale scores are logit scores exported from Winsteps, while the other scales' scores are sum scores calculated after reverse coding any items as appropriate.

Table 36, on the following page, contains the correlation matrix with the three dimensions of the Comparative Mathematics Anxiety Scale and all of the other mathematics constructs described above. Note the strong, positive correlations among the three subscales—0.778 between the emotional-attitudinal and mental-cognitive subscales, 0.831 between the emotional-attitudinal and physical-somatic subscales, and 0.725 between the mental-cognitive and physical-somatic subscales. The three subscales have correlations with the mAMAS of 0.650, 0.655, and 0.669, respectively. As desired, they are relatively strong correlations since they all measure mathematics anxiety, but the correlations are weaker than the correlations among the subscales, which is to be expected. As final evidence for the convergent validity of the subscales, they have moderate, positive correlations of 0.355, 0.331, 0.334 with the fixed mindset scale. Finally, providing divergent validity evidence, there are moderate to moderately-strong negative correlations between the three subscales and the following mathematics scales: confidence (-0.620, -0.560, -0.588), motivation (-0.543, -0.451, -0.498), perseverance (-0.442, -0.448, -0.401), and growth mindset (-0.379, -0.367, -0.409). In summary, there is ample evidence of the criterion validity for the three dimensions of the Comparative Mathematics Anxiety Scale.

Table 36

Comparative Math Anxiety Scale (CMAS), Criterion Validity Analysis Correlation Matrix

Scale	*mAMAS*	*conf.*	*motiv.*	*persev.*	*growth*	*fixed*	*A1*	*A2*	*A3*
mAMAS	1	-.664	-.580	-.442	-.424	.454	.650	.655	.669
conf.		1	.819	.587	.557	-.515	-.620	-.560	-.588
motiv.			1	.600	.495	-.461	-.543	-.451	-.498
persev.				1	.590	-.458	-.442	-.448	-.401
growth					1	-.628	-.379	-.367	-.409
fixed						1	.355	.331	.334
A1							1	.778	.831
A2								1	.725
A3									1

Note. All correlations were significant at the 0.01 level (1-tailed test). A1 = emotional-attitudinal dimension of the CMAS, A2 = mental-cognitive dimension of the CMAS, and A3 = physical-somatic dimension of the CMAS.

To further understand the relationship between the three dimensions of the CMAS, we can look at a scatterplot for each pair of dimensions (see *Figures 40-42*, on the following pages). Note that the choice of the predictor (or the variable on the x-axis) was arbitrary in each scatterplot, as no hypotheses were being made about the directionality of the predictive relationship between the subscales of the CMAS.

Figure 40

Scatterplot of A1 Emotional-Attitudinal and A2 Mental-Cognitive Scores

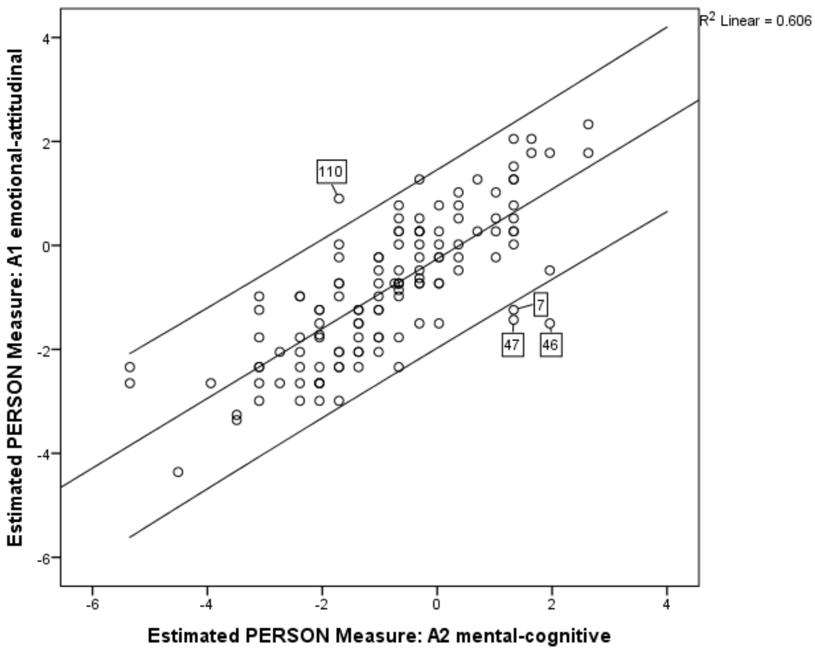


Figure 41

Scatterplot of A1 Emotional-Attitudinal and A3 Physical-Somatic Scores

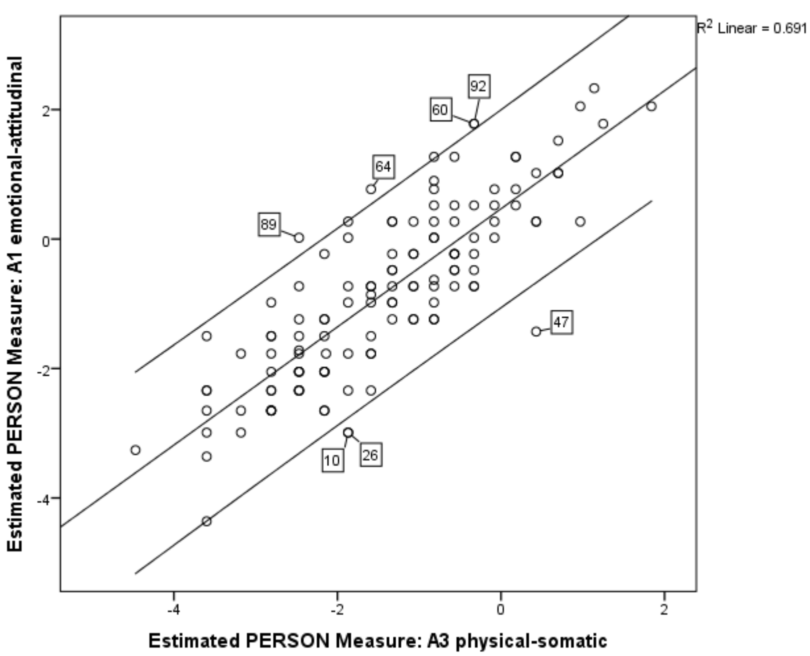
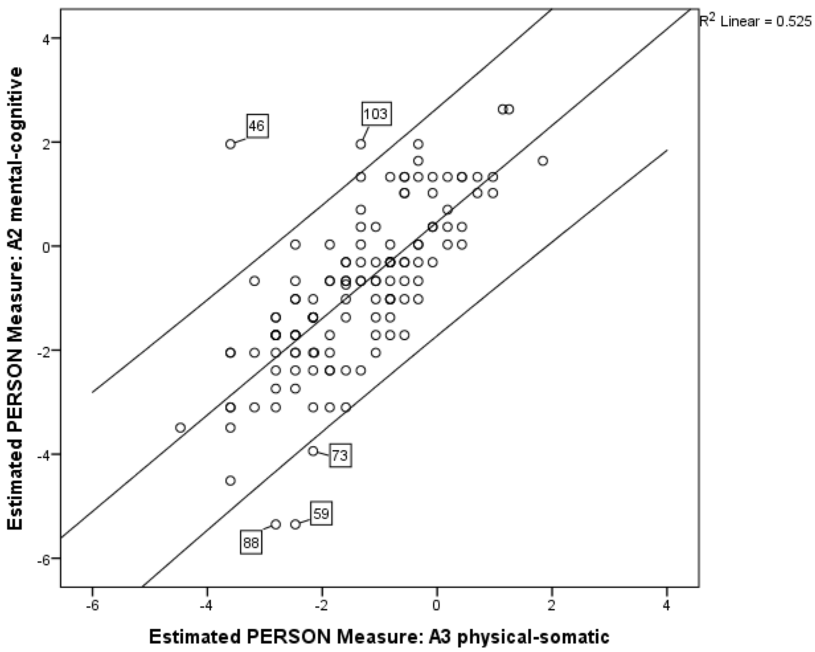


Figure 42

Scatterplot of A2 Mental-Cognitive and A3 Physical-Somatic scores



Across all three figures we see a very strong, positive relationship between each of the dimensions. This supports the evidence shown by the strong, positive correlations between the subscales. That being said, in each figure, students have been highlighted that are furthest away from the linear fit line, indicating their scores on the two dimensions being observed are not so perfectly related. For example, student 46 in *Figure 40* has an emotional-attitudinal logit score around -2 but a mental-cognitive logit score around +2. This same student is shown to have a physical-somatic logit score around -4 in *Figure 42*. These figures provide evidence that while the subscales may be highly correlated, the dimensions are unique constructs along which student scores can vary in unexpected ways.

CHAPTER 5 – DISCUSSION

The previous chapter presented the results of this dissertation research, which confirmed that a Rasch-based scale to measure the three dimensions of mathematics anxiety can be done in a reliable, valid, and meaningful way. In this chapter, I discuss this finding in more detail and how this study contributes to the existing body of research on developing instruments to measure mathematics anxiety, particularly the emotional-attitudinal, mental-cognitive, and physical-somatic dimensions of the construct. The chapter concludes with a discussion of the limitations of the study, areas for future research, as well as implications and conclusions.

Revisiting the Overarching Research Question

The overarching research question for this study was: *To what extent can the construct of mathematics anxiety be measured reliably and meaningfully by developing a Rasch-based scale?* This section will provide an overview of the findings related to this research question, including a discussion of how these findings relate to what is already known in the field about the construct of mathematics anxiety and about measuring mathematics anxiety.

Overview of Findings

This research question was answered by employing Rasch measurement theory (Andrich, 1978; Ludlow et al., 2014; Rasch, 1960/1980; Wright & Masters, 1982) and Rasch rating scale analysis (Wright & Masters, 1982) to develop and refine a scale measuring three dimensions of mathematics anxiety: emotional-attitudinal, mental-cognitive, and physical-somatic. In addition, cognitive interviews (Willis, 1999) were conducted and analyzed qualitatively to further understand the construct and refine the instrument. Rasch rating scale analysis was used to analyze the data from all three rounds of survey data collection. In the first round, evidence for the concept of the scale (comparative response options and third-person item statements) was found, as well as some preliminary evidence to support the conceptual framework. The second round of data collection provided some more evidence that supported the revisions to the scale, but it also raised some unique challenges and questions about the item format and its interpretability for a broad range of high school students. After the second round of survey data collection and Rasch rating scale analysis, cognitive interviews were conducted to gather further information from students about how they interpreted the scale instructions, the wording of the items, and the comparative response options. Cognitive interviews were also conducted for a set of students where they completed a card-sorting activity to further flesh out the conceptualization of the three dimensions of mathematics anxiety.

The third and final round of data collection led to a three-dimensional scale containing 23 third-person items that can be used to measure high school students' levels of mathematics anxiety across three dimensions in a reliable, valid, and meaning way (see *Figures 31-39* for the construct definition map and *Tables 33-35* for the interpretations of the raw scale scores for each dimension). The findings suggest that the empirical data fit the Rasch model and *some* of the Rasch measurement principles (e.g., variation). Relatedly, the scale partially confirms the *a priori* theory and construct definition of mathematics anxiety. That is, there are three unique but highly-correlated dimensions of mathematics anxiety and there are levels of anxiety within each of these dimensions. However, more work is needed to fully conceptualize the ladder-like progression from lower to higher levels of mathematics anxiety.

In addition to developing and refining both the tripartite conceptual framework for mathematics anxiety and the Comparative Mathematics Anxiety Scale (CMAS), data was also collected to assess the measurement invariance and the validity of the scale. Overall, the scale shows fairly strong measurement invariance across grade and gender, with weaker measurement invariance for self-reported English-language proficiency (ELP). Specifically, no significant differential item functioning (DIF) was identified for grade across all three dimensions, and only one significant difference was found for gender. There is no clear reason or hypothesis for this gender DIF. The mental-cognitive dimension of mathematics anxiety showed measurement invariance for ELP; however, significant DIF for ELP was found for three items on the emotional-attitudinal dimension and two items on the physical-somatic dimension of mathematics anxiety. Items showing DIF were some of the hardest and easiest items, meaning students who are not ELP may have more trouble interpreting and responding to the items on the CMAS. Construct validity evidence was already discussed as it relates to the interpretability of the scale scores. Finally, the criterion validity evidence for the scale is extremely strong. The three dimensions of the CMAS are most highly and positively correlated to the other dimensions of the CMAS and have strong positive correlations with the mAMAS, moderate positive correlations with the fixed mindset scale, and negative correlations with the other constructs. In addition, the scatterplots provide some evidence that while the three dimensions of CMAS are very highly correlated, students can vary across the three dimensions. Therefore, the three dimensions of mathematics anxiety are unique constructs from a measurement perspective.

Discussion of Findings

This dissertation contributes to the literature in several ways. First, it expands upon and confirms some of what is already known about the conceptualization of mathematics anxiety. Specifically, there are multiple dimensions to mathematics anxiety and each dimension can span from very low levels to very high levels of anxiety. In addition, students can vary in their level of anxiety across the three dimensions; that is, scoring high on one dimension of the scale does not necessarily mean that a student will score high on another dimension. This work confirms that mathematics anxiety includes at least three dimensions: emotional-attitudinal, mental-cognitive, and physical-somatic, most directly building off of the works of Wigfield and Meece (1988) and Cavanagh and Sparrow (2011). It expands the work on these dimensions by clearly articulating how those dimensions can be defined conceptually and by showing that all of the dimensions can and should span all levels of mathematics anxiety.

Specifically, as shown in *Figures 31-33* and *Table 33*, emotional-attitudinal anxiety can be clearly defined from low to high mathematics anxiety. Students with low emotional-attitudinal mathematics anxiety are unfazed and in their element when doing mathematics. Moving up to moderate anxiety, students may begin to feel a little intimated by the mathematics and have a sense of worry when doing mathematics. Continuing up the progress, students with high emotional-attitudinal mathematics anxiety feel a bit apprehensive, emotionally worked up, and may begin to feel distressed. Students with very high emotional-attitudinal anxiety are terrified of mathematics and mathematics problems.

Mental-cognitive mathematics anxiety can be similarly described as seen in *Figures 34-36* and *Table 34*. At the low end, students can think clearly and are extremely focused when doing mathematics. As students' level of anxiety begins to increase, students' thoughts may become scattered and they feel confused when doing mathematics. Students with high mental-cognitive mathematics anxiety can't recall what they learned when working on mathematics problems and their mind keeps blanking out. At the top, those with extremely high mental-cognitive mathematics anxiety can think of nothing else except their fear of mathematics when presented with mathematics problems.

Finally, *Figures 37-39* and *Table 35* explicate the physical-somatic mathematics anxiety dimension. Students with low physical-somatic mathematics anxiety feel physically at ease and calm. As symptoms of physical-somatic mathematics anxiety begin to emerge, students with moderate physical-somatic mathematics anxiety feel a little tense when doing mathematics, while those with moderately-high anxiety may begin to develop a headache when doing mathematics. At the high level of physical-somatic mathematics anxiety, students' palms may start to sweat, their heart may be racing, and they may feel physically uncomfortable. Students that feel nauseous to the point of feeling like they are going to be sick when doing mathematics have extremely high physical-somatic anxiety.

Given the empirical data fit the Rasch rating scale model fairly well, the third-person items provide a meaningful interpretation of one's level of mathematics anxiety across all three dimensions. Again, as presented in *Tables 33-35*, a student's raw score has a corresponding description of that student's level of mathematics anxiety at the time they completed the survey. The score descriptors allow for the diagnosis of a student's level of mathematic anxiety for each dimension, which is a unique benefit for instruments developed with Rasch measurement theory. As *Figures 40-42* showed, students may score high on one dimension of anxiety but not another, which is evidence of the unique effect the dimensions have on a person. This suggests that students should be represented by scores for all three dimensions as a type of score portfolio. Depending on where a student scores across each of the three dimensions, different interventions may be appropriate. That is, an intervention that is appropriate for a student with high mental-cognitive anxiety but low emotional-attitudinal and physical-somatic anxiety may not be the appropriate intervention for a student that has high anxiety across all three dimensions.

An additional contribution to the field is that this work employed Rasch measurement principles in the *development* of the scale and did not just apply Rasch rating scale analysis without the theoretical underpinnings. By employing Rasch measurement and scale development principles, this work clearly defined the target rather than basing decisions solely on statistical results. This appears to be the first scale development process in the mathematics anxiety literature to have taken this approach from the beginning in a clear, intentional way.

Similarly, this work improves upon many of the prior instruments by employing Rasch measurement theory and rating scale analysis over classical test theory methods of analysis, resulting in greater construct validity evidence, greater ability to identify items and persons that do not fit the model, and greater interpretability and usability of scores obtained from the scale (Wright & Masters, 1982). First, because CTT focuses on items as replicates of the construct where items are ideally highly correlated and all have relatively moderate levels of difficulty, there is little in the way of validating the definition of the construct. This is particularly true in contrast to Rasch measurement theory and rating scale analysis where items lie along a hierarchical continuum of varying construct difficulty and the variable map acts as a direct way to compare the results to the theorized construct definition. Therefore, instruments such as the CMAS that are developed using Rasch measurement theory and rating scale analysis have stronger construct validity evidence than scales developed under CTT, which are most of the mathematics anxiety scales.

Second, Rasch measurement theory employs probabilistic models in its item and person estimation and CTT does not. Therefore, Rasch measurement theory has the advantage of being able to calculate an expected score for a person on an item which can then be compared to the observed score to determine how well the model fit the data and assess the fit of the person and item statistics. This makes it possible to identify individual items that need to be revised and specific persons that may have had trouble responding to the items as intended. This is a huge advantage of Rasch measurement theory over CTT, especially in instrument development. This was seen throughout the CMAS development process and item and person misfit was analyzed and used to make decisions about survey and item revisions.

The use of probabilistic models under Rasch measurement theory also allows for the transformation of raw scores into logit scores, which is an interval measurement scale (Wright & Masters, 1982). This means logits are linear such that a one-logit difference between persons and items has the same probability of an item being answered correctly regardless of where you are looking at on the scale. This is not true of methods under CTT which rely on raw scores. This makes scales developed with Rasch measurement theory more useful and interpretable.

Finally, another aspect of Rasch measurement theory that makes scores more useful and interpretable is the fact that item and person estimates are person-free and item-free, respectively (Wright & Masters, 1982). In contrast, item and person estimates from CTT methods are sample dependent and test/scale dependent, respectively, meaning it is challenging to compare performance on a test or scale across groups or across time points. This is perhaps the greatest advantage of scales developed with Rasch measurement theory methods over CTT methods—information about individual persons and specific items is applicable outside of the context of the instrument and the sample.

Limitations

There are several limitations, or potential limitations, of this research. Specifically, study limitations include: the size and representativeness of the sample, variations in survey formatting and administration, the uniqueness of the item format, the lack of access to survey respondents, the somewhat limited amount of validity evidence, and possibly the analytic method chosen. Each limitation will be described in some detail below.

Sample Size and Representativeness

The sample size for the first and third rounds of data collection were relatively small at 77 and 132 students, respectively, both of which are very close to the minimum recommendation of 5-10 times the number of respondents as the number of items. In addition, both rounds of data came from a single high school. Therefore, it is hard to make claims that these samples were representative of the larger population of high school students in the United States. For the second round of data collection, while the sample sizes were varied for the three dimensions, they were quite a bit larger. However, no demographics data was reported to indicate the representativeness of those who completed the surveys. For the third and final round of data collection, juniors and female students were overrepresented in the sample, and there were only twelve students who identified as non-English language proficient. Other demographic data was not collected, so it is unclear how representative the sample was on other student characteristics.

Variability in Survey Format and Administration

Each round of data collection varied in the format and/or mode of survey administration. In the first round of data collection, all anxiety items were together on a single survey, clearly formatted on the instrument, and administered using a paper-and-pencil survey and Scantron answer sheet. The second round of data collection was similarly administered using a paper-and-pencil survey and Scantron answer sheet. However, because the items were being included on a survey begin used as part of a larger study with many other important scales, each dimension of the CMAS was administered on a different survey form (A-C) and the items were not as clearly formatted on the instrument. Finally, the third round of data collection was more similar to the first round in that all anxiety items were together on a single survey and were clearly formatted, but the instrument was administered electronically via Qualtrics because of COVID-19. Because of the varying modes of administration and scale formatting, it is a little more difficult to parse out the reasons for the changes and challenges. That is, are changes due to item revisions or to the format or mode of scale and survey administration? It is likely that both are at play.

Unfamiliarity with the Item Format

In the field of measurement, third-person items with comparative response options are a very new and unfamiliar item format, especially for this population of high school students. With lack of familiarity, one is required to put in more cognitive effort to understand the task and what is being asked of you. This requires a level of focus and engagement with the survey that some students may not have had. In addition, even with enough focus the items and response options may have simply been too different for some students to understand and appropriately use. The cognitive interviews provide some evidence to both the focus and the understanding argument. In addition, the cluster analysis from the second round of data collection as well as the misfit statistics from the first and third round given credence to the possibility of having two groups of respondents—those that read and understand the directions and items and those who did not.

Access to Survey Respondents

For all first and third rounds of data collection, the data was collected anonymously, and for the second, the data was collected as part of a larger study where later access to students was not permissible. Because of this, it was impossible to interview students who completed the survey and scale items to better understand their thought process, especially as it is related to unexpected responses or person and item misfit. This limitation was somewhat remedied by re-administering the second version of the survey to a small group of students and conducting cognitive interviews with them. While this data was invaluable in the interpretation of results and further revision of the scale, it was not a perfect replacement for interviews with students that were part of the rounds of survey data collection that were analyzed and presented here.

Limited Validity Evidence

While the validity evidence presented in this work is fairly strong, the research was somewhat limited in the type of validity evidence available. As stated earlier, very few demographic variables were collected from students. The most notable missing student demographic is race. This was in part due to small sample size issues and wanting to reduce the risk of students being identifiable based on the demographic data collected but also due to the sensitivity of asking this question. In addition, validity evidence for outcome data could not be collected and analyzed as part of this study. For example, it would have been relevant to collect students' mathematics grades and assessment data, as well as data on the mathematics course they are in or the courses they have taken, as well as their future plans in mathematics. Finally, data for individual participants was only collected at a single time point, therefore we are unable to look at the sensitivity of the scales to measuring change in the construct (e.g., pre/post intervention). These types of data would increase the validity evidence for the CMAS.

Analyzing the Multidimensionality of the Construct

One aspect of the study that could also be considered a limitation is the choice of analytic method. For this work, I ran three separate unidimensional Rasch analyses for the three unique dimensions of mathematics anxiety, what some call the consecutive approach (Briggs & Wilson, 2003). However, some may consider the three dimensions of mathematics anxiety to be a single construct that also happens to be multidimensional, where the appropriate analytic method would have been multidimensional Rasch analysis. There are advantages to using multidimensional Rasch analysis, including increased reliability in the estimates, increased accuracy in estimating the correlations among the dimensions, and better statistical fit of the model to the data (Briggs & Wilson, 2003). However, these are all statistical advantages that do not take into account the conceptualization of the construct. While emotional-attitudinal, mental-cognitive, and physical-somatic are all dimensions of mathematics anxiety and known to be highly correlated, in this study they were conceptualized as unique constructs that could have different relationships to other constructs of interest, such as mathematics grades and assessment scores. If multidimensional Rasch analysis had been used, the estimations for each dimension would have been inextricably linked. It was important for this research to present how each dimension behaved separately, unaffected by the other dimensions.

Implications

The development of the tripartite CMAS has implications for future research on mathematics anxiety, as well as for the field of measurement. With the increase of technology over the past few decades, the number of STEM positions has increased substantially and the need for mathematical competency and performance has increased along with it. Unfortunately, mathematics anxiety has been found to be a prevalent and significant barrier to increased mathematics achievement (Ashcraft & Moore, 2009). And it may be a more significant barrier for some subgroups of students (e.g., girls). As we work to diversify the STEM workforce, it will be critical to reduce the impact of any construct found to negatively influence one's future interest in STEM, including college major and career course. Therefore, being able to measure mathematics anxiety in a valid, reliable, and meaningful way is important for the advancement and diversification of mathematics and STEM education broadly.

Specifically, this scale allows for the measurement of three dimensions of mathematics anxiety, that while highly correlated, may interact with mathematics performance and future plans for mathematics in different ways. By measuring mathematics anxiety along these three dimensions, we can determine how mathematics anxiety impacts students' mathematics performance and participation in mathematics activities, courses, college majors, and careers, and whether these impacts vary for different subgroups of students. This is critical for diagnosing student mathematics anxiety and developing interventions to retain students in mathematics and other STEM fields.

The CMAS and corresponding figures (*Figures 31-39)* and tables (*Tables 33-35*) have the ability to assess students' levels of mathematics anxiety, take the resulting students' scores, and diagnose their level of mathematics anxiety across the three dimensions. In this way, the CMAS and the construct validity figures and tables act as a critical diagnostic tool in the assessment and selection of students for interventions. Given the multiple dimensions of mathematics anxiety and students' ability to vary across the three dimensions, interventions could and should be developed that are attuned to these nuances and target the areas of mathematics anxiety where students are most at risk. Then, *Tables 33-35* can be used to diagnose students and place them into an appropriate intervention or set of interventions. Finally, this measure also provides a way to assess the impact of such interventions on student outcomes of interest.

In addition to impacting research on mathematics anxiety, the development of the CMAS has implications for the field of measurement as well. By following a clear development process and relying on Rasch measurement theory to clearly articulate the construct of interest and how to assess and refine it, this study has shown the promise of Rasch measurement theory as a guide for the development of reliable, valid, and meaningful instruments. This study has also shown the potential promise, but also challenge, of developing third-person items with comparative response options. With a stronger theoretical definition for the construct of mathematics anxiety, I am confident the CMAS could be improved to more accurately measure the emotional-attitudinal, mental-cognitive, and physical-somatic dimensions of mathematics anxiety.

Areas for Future Research

Based on the work done here and the implications discussed above, there are several areas that call for future research. Impacting both the study of mathematics anxiety and the future measurement of the construct, a more accurate, clear definition of the construct of mathematics anxiety is critical. Convening a panel of mathematics education and psychology experts to further flesh out and wrestle with how to define mathematics anxiety along these three, and potentially other, dimensions is critical for moving this work forward. Having a clear, consistent definition of the construct is important both for having a common language in which to talk about what mathematics anxiety is but also for studying the construct. Without a stronger conceptual definition of the construct, the other recommendations here might not be so fruitful.

Second, the CMAS requires further refinement to better capture the entire continuum of mathematics anxiety from the very lowest end to the very highest end, with the additional goal of having the item locations for the three subscales stabilize across administrations. While the current study has done much to advance the development of an instrument that measures the three dimensions of emotional-attitudinal, mental-cognitive, and physical-somatic mathematics anxiety, there are still places where the scale could use additional work. The first focus for continued instrument development should be on this—expanding the items along the entire length of the continuum and stabilizing the item locations.

Relatedly, it will be important to test the instrument with additional, varied populations of students during this instrument revision process. While getting the items to extend the entire length of the continuum and stabilize across administrations is important, it is equally important that this is the case for the entire population of high school students. Therefore, it is important that a diverse sample of students is used to test survey revisions. Otherwise, we have not truly developed an instrument that is valid for the entire population of interest. Of particular note, it is critical to assess the use of the CMAS with students who do not identify as English-language proficient, as well saw some evidence that the instrument may perform differently for these students. Whether or not the instrument is appropriate for that population of students is yet to be determined and could use focused attention.

Once the construct of mathematics anxiety has been more clearly defined and the CMAS has been revised and stabilized to extend the entire continuum of mathematics anxiety, it would be of additional interest to pilot additional item formats. While the third-person items and comparative response options have some benefits in defining the construct and placing students on the construct, there is some evidence that not all students respond to the item format as expected. Therefore, it may be worthwhile to explore other formats. One that is of particular interest is having the items on the CMAS listed for students and asking them to select the item or items that sound like themselves or align with their experience of mathematics anxiety. Based on which item or items a student selects, we can determine their level of anxiety based on where the item(s) are located on the variable maps. That is, the work of developing the construct using Rasch measurement theory and third-person items with comparative response options would lay the groundwork for being able to ask the question in a more direct, potentially clearer way.

Finally, conducting additional validity studies on the construct of mathematics anxiety and the CMAS are vital for moving the work forward and providing additional evidence for the utility of the measure. Of particular importance is determining the instruments' sensitivity and ability to measure change over time. With appropriate validity evidence, the CMAS could improve the diagnosis of mathematics anxiety and in measuring the impact of interventions of reducing, hopefully, students' experience of mathematics anxiety.

REFERENCES

Adams, R. J., Wilson, M., & Wang, W. (1997). The multidimensional random coefficients multinomial logit model. *Applied Psychological Measurement, 21*(1), 1-23.

Adams, R. J., Wu, M. L., & Wilson, M. R. (2015). ACER ConQuest: Generalised Item Response Modelling Software (Version 4.0.) [Computer program]. Camberwell, Victoria: Australian Council for Educational Research.

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*, 143-150.

Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika, 43,* 561-573.

Ashcraft, M. H., Kirk, E. P., & Hopko, D. (1998). "On the cognitive consequences of mathematics anxiety," in *The Development of Mathematical Skills,* ed. C. Donlan (Hove: Erlbaum), 175-196.

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

Ashcraft, M. H., & Kirk, E. P. (2001). The relationships among working memory, math anxiety, and performance. *Journal of Experimental Psychology: General, 130*, 224-237.

Ashcraft, M. H., & Moore, A.M. (2009). Mathematics anxiety and the affective drop in performance. *Journal of Psychoeducational Assessment, 27*(3), 197-205.

Baloglu, M. & Koçak, R. (2006). A multivariate investigation of the differences in mathematics anxiety. *Personality & Individual Differences, 40*(7), 1325-1335.

Baloglu, M. & Zelhart, P. F. (2007). Psychometric properties of the revised mathematics anxiety rating scale. *Psychological Record, 57*(4), 593-612.

Bandalos, D. L., Yates, K., & Thorndike-Christ, T. (1995). Effects of math self-concept, perceived self-efficacy, and attributes for failure and success on test anxiety. *Journal of Educational Psychology, 87,* 611-623.

Beasley, T. M., Long, J.D., & Natali, M. (2001). A confirmatory factor analysis of the Mathematics Anxiety Scale for Children. *Measurement & Evaluation in Counseling & Development, 34*, 14-26.

Beilock, S. L., Gunderson, E. A., Ramirez, G., & Levine, S. C. (2010). Female teachers' math anxiety affects girls' math achievement. *Proceedings of the National Academy of Science, USA, 107*, 1860-1863.

Beilock, S. L., Rydell, R. J., & McConnell, A. R. (2007). Stereotype threat and working memory: Mechanism, alleviations, and spillover. *Journal of Experimental Psychology: General, 136*, 256-276.

Betz, N. (1978). Prevalence, distribution, and correlates of math anxiety in college students. *Journal of Counseling Psychology, 25*, 441-448.

Blatchford, P. (1996). Pupils' views on school work and school from 7 to 16 years. *Research Papers in Education, 11*, 263-288.

Briggs, D. C., & Wilson, M. (2003). An introduction to multidimensional measurement using Rasch models. *Journal of Applied Measurement, 4*(1), 87-100.

Brown, M., Brown, P., & Bibby, T. (2008). "I would rather die": Reasons given by 16-year-olds for not continuing their study of mathematics. *Journal for Research in Mathematics Education, 10*, 3-18.

Brush, L. (1981). Some thoughts for teachers on mathematics anxiety. *Arithmetic Teacher, 29*(4), 37-39.

Carey, E., Hill, F., Devine, A., and Szűcs, D. (2017). The Modified Abbreviated Math Anxiety Scale: A valid and reliable instrument for use with children. *Frontiers in Psychology, 8*(11), 1-13. doi: 10.3389/fpsyg.2017.00011

Catsambis, S. (1994). The path to math: Gender and racial-ethnic differences in mathematics participation from middle school to high school. *Social Education, 67*, 199-215.

Cavanagh, R., & Sparrow, L. (2011). Mathematics anxiety: Scaffolding a new construct model. *Mathematics Traditions and [New]Practices*. In Proceedings of the Annual Conference of Mathematics Education Research Group of Australasia, Alice Springs.

Cheng, Y., Wang, W., & Ho, Y. (2009). Multidimensional Rasch analysis of a psychological test with multiple subtests: A statistical solution for the bandwidth-fidelity dilemma. *Educational and Psychological Measurement, 69*(3), 369-388.

Chinn, S. (2009). Mathematics anxiety in secondary students in England. *Dyslexia, 15*, 61-68.

Chipman, S. F., Krantz, D. H., & Silver, R. (1992). Mathematics anxiety and science careers among able college women. *Psychological Science, 3*, 292-295.

Chou, Y.-T., & Wang, W.-C. (2010). Checking dimensionality in item response models with principal component analysis on standardized residuals. *Educational and Psychological Measurement, 70*(5), 717–731.

Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment, 7*(3), 309–319. http://dx.doi.org/ 10.1037/1040-3590.7.3.309.

Cooper, S. E., & Robinson, D. A. (1991). The relationship of mathematics self-efficacy beliefs to mathematics anxiety and performance. *Measurement and Evaluation in Counseling and Development, 24*, 4-11.

Crocker, L., & Algina, J. (1986). *Introduction to classical & modern test theory*. Mason, OH: Cengage Learning.

CTB/McGraw-Hill. (2007a). *Acuity Algebra Readiness Technical Report*. Monterey, CA: Author.

CTB/McGraw-Hill. (2007b). *Acuity Algebra Proficiency Technical Report*. Monterey, CA: Author.

D'Ailly, H., & Bergering, A. J. (1992). Mathematics anxiety and mathematics avoidance behaviour: A validity study of two factors. *Educational and Psychological Measurement, 52*(2), 369-378.

Deffenbacher, J. L. (1980). Worry and emotionality in test anxiety. In I. G. Sarason (Ed.), *Test anxiety: Theory, research, and applications* (pp. 111-128). Hillsdale, NJ: Erlbaum.

DeVellis, R. F. (2003). *Scale development: Theory and applications* (2nd ed.). Newbury Park: Sage Publications.

Devine, A., Fawcett, K., Szucs, D., & Dowker, A. (2012). Gender differences in mathematics anxiety and the relation to mathematics performance while controlling for test anxiety. *Behavioral and Brain Functions, 8*, 1-9.

Dew, K. M. H., Galassi, J. P., & Galassi, M. D. (1983). Mathematics anxiety: Some basic issues. *Journal of Counseling Psychology, 30,* 443-446.

Dowker, A. D. (2005). *Individual Differences in Arithmetic: Implications for Psychology Neuroscience and Education*. Hove: Psychology Press.

Dowker, A., Sarkar, A., & Looi, C. Y. (2016). Mathematics anxiety: What have we learned in 60 years? *Frontiers in Psychology, 7*, Article 508, 1-16.

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

Else-Quest, N., Hyde, J. S., & Linn, M. (2010). Cross-national patterns of gender differences in mathematics: A meta-analysis. *Psychological Bulletin, 136*, 103-127.

Eysenck, M.W., & Calvo, M. G. (1992). Anxiety and performance: The processing efficiency theory. *Cognition and Emotion, 6*, 409-434.

Faust, M. W. (1992). *Analysis of physiological reactivity in mathematics anxiety*. Unpublished doctoral dissertation, Bowling Green State University, Bowling Green, OH.

Fennema, E., & Sherman, J. A. (1976). Fennema-Sherman mathematics attitudes scales: Instruments designed to measure attitudes towards the learning of mathematics by females and males. *Journal for Research in Mathematics Education, 7*(5), 324-326.

Ferrando, P. J., Varea, M. D., & Lorenzo, U. (1999). Evaluación psicométrica del cuestionario de ansiedad y rendimiento (CAR) en una muestra de escolares. *Psicothema, 11*, 225-236.

Furner, J. & Berman, B. (2003). Math anxiety: Overcoming a major obstacle to the improvement of student math performance. *Childhood Education, 79*(3), 170-174.

Goetz, T., Cronjaeger, H., Frenzei, A. C., Ludtke, O., & Hall, L. C. (2010). Academic self-concept and emotion relations: Domain specificity and age effects. *Contemporary Educational Psychology, 35*, 44-58.

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate data analysis* (7th ed.). Upper Saddle River, NJ: Pearson.

Hambleton, R., & Jones, R. (1993). An NCME instructional module on comparison of classical test theory and item response theory and their application to test development. *Educational Measurement: Issues and Practice, 12*(3), 38-47.

Hembree, R. (1988). Correlates, causes, effects, and treatment of test anxiety. *Review of Educational Research, 58*, 47-77.

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

Ho, H., Senturk, D., Lam, A. G., Zimmer, J. M., Hong, S., Okamoto, Y., et al. (2000). The affective and cognitive dimensions of math anxiety: A cross-national study. *Journal for Research in Mathematics Education, 31*, 362-379.

Hoffman, B. (2010). 'I think I can, but I'm afraid to try': The role of self-efficacy beliefs and mathematics anxiety in mathematics problem solving efficiency. *Learning and Individual Differences,* 20, 276-283.

Holland, P., & Wainer, H. (Eds.). (1993). *Differential item functioning*. New York: Routledge.

Hopko, D. R. (2003). Confirmatory factor analysis of the Math Anxiety Rating Scale—Revised. *Educational and Psychological Measurement, 63*, 336-351.

Hopko, D. R., Ashcraft, M.H., Gute, J., Ruggiero, K. J., & Lewis, C. (1998). Mathematics anxiety and working memory: Support for the existence of a deficient inhibition mechanism. *Journal of Anxiety Disorders, 12*, 343-355.

Hopko, D. R., Mahadevan, R., Bare, R. L., & Hunt, M. A. (2003). The Abbreviated Math Anxiety Scale (AMAS): Construction, validity, and reliability. *Assessment, 10*, 178-182.

Hopko, D. R., McNeil, D., W., Lejuez, C. W., Ashcraft, M. H., Eifert, G. H., & Reil, J. (2003). The effects of anxious responding on mental arithmetic and lexical decision task performance. *Journal of Anxiety Disorders, 17*, 647-665.

Hyde, J. S. (2005). The gender similarities hypothesis. *American Psychologist, 60*, 581-592.

Hyde, J. S., Fennema, E., Ryan, M., Frost, L., & Hopp, C. (1990). Gender comparison of mathematics attitudes and affect: A meta-analysis. *Psychology of Women Quarterly, 14*, 299-324.

IBM Corporation [IBM]. (2013). IBM SPSS Statistics for Windows (Version 22.0.) [Computer program]. Armonk, NY: IBM Corporation.

Jackson, C. & Leffingwell, R. (1999). The role of instructors in creating math anxiety in students from kindergarten through college. *Mathematics Teacher, 92*(7), 583-587.

Jadjewski, K. (2011). *Attitudes about Mathematics: Compare and Contrast Boys and Girls from High and Low Socio-economic Status*. Masters thesis, California State University.

Jain, S. & Dowson, M. (2009). Mathematics anxiety as a function of multidimensional self-regulation and self-efficacy. *Contemporary Educational Psychology, 34*(3), 240-249.

Johns, M., Schmader, T., & Martens, A. (2005). Knowing is half the battle: Teaching stereotype threat as a means of improving women's math performance. *Psychological Science, 16*, 175-179.

Johnston-Wilder, S., Brindley, J., & Dent, P. (2014). *Technical Report: A Survey of Mathematics Anxiety and Mathematical Resilience amongst Existing Apprentices*. Coventry: University of Warwick.

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

Kellogg, J. S., Hopko, D. R., & Ashcraft, M. H. (1999). The effects of time pressure on arithmetic performance. *Journal of Anxiety Disorders, 13*, 591-600.

Kena, G., Hussar W., McFarland J., de Brey C., Musu-Gillette, L., Wang, X., Zhang, J., Rathbun, A., Wilkinson-Flicker, S., Diliberti M., Barmer, A., Bullock Mann, F., and Dunlop Velez, E. (2016). *The Condition of Education 2016* (NCES 2016-144). U.S. Department of Education, National Center for Education Statistics. Washington, DC.

Lang, P. J. (1968). Fear reduction and fear behavior: Problems in treating a construct. In J. M. Schlien (Ed.), *Research in psychotherapy* (Vol. III). Washington, D.C.: American Psychological Association.

Lee, J. (2009). Universals and specifics of math self-concept, math self-efficacy, and math anxiety across 41 PISA 2003 participating countries. *Learning and Individual Differences,* 19, 355-365.

Liebert, R., & Morris, L. (1967). Cognitive and emotional components of test anxiety: A distinction and some initial data. *Psychological Reports, 20*, 975-978.

Linacre, J. M. (n.d.). *DIF - DPF - bias - interactions concepts*. Retrieved February 20, 2021, from https://www.winsteps.com/winman/difconcepts.htm

Linacre, J. M. (1998a). Detecting multidimensionality: Which residual data-type works best? *Journal of Outcome Measurement, 2*, 266-283.

Linacre, J. M. (1998b). Structure in Rasch residuals: Why principal components analysis? *Rasch Measurement Transactions, 12*, 636.

Linacre, J. M. (2016). Winsteps® Rasch measurement computer program User's Guide. Beaverton, Oregon: Winsteps.com.

Linacre, J. M. (2017). WINSTEPS (Version 4.0.1.) [Computer program]. Beaverton, OR: Winsteps.com.

Liu, O. L., Wilson, M., & Paek, I. (2008). A multidimensional Rasch analysis of gender differences in PISA mathematics. *Journal of Applied Measurement, 9*(1), 18-35.

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.

Louie, J., Rhoads, C., & Mark, J. (2016). Challenges to using the regression discontinuity design in educational evaluations: Lessons from the Transition to Algebra Study. *American Journal of Evaluation, 37*(3), 381–407. https://doi.org/10.1177/1098214015621787

Lubienski, S. T. (2002). *Are we achieving "mathematical power for all?" A decade of national data on instruction and achievement*. Paper presented at the meeting of the American Educational Research Association, New Orleans: LA.

Ludlow, L. H. (1983). *The analysis of Rasch model residuals* [Unpublished doctoral dissertation]. University of Chicago.

Ludlow, L. H. (1986). Graphical analysis of item response theory residuals. *Applied Psychological Measurement, 10*(3), 217-229.

Ludlow, L. H., & Haley, S. M. (1995). Rasch model logits: Interpretation, use, and transformation. *Educational and Psychological Measurement, 55*(6), 967-975.

Ludlow, L. H., Matz-Costa, C., Johnson, C., Brown, M., Besen, E., & James, J. B. (2014). Measuring engagement in later life activities: Rasch-based scenario scales for work, caregiving, informal helping, and volunteering. *Measurement and Evaluation in Counseling and Development*. 47(2), 127-149.

Ludlow, L. H., Matz-Costa, C., & Klein, K. (2019). Enhancement and validation of the Productive Engagement Portfolio–Scenario (PEP–S8) scales. *Measurement and Evaluation in Counseling and Development, 52*(1), 15-37.

Ma, X. (1999). A meta-analysis of the relationship between anxiety toward mathematics and achievement in mathematics. *Journal for Research in Mathematics Education, 30*, 520-541.

Ma, X. & Kishor, N. (1997). Assessing the relationship between attitude toward mathematics and achievement in mathematics: A meta-analysis. *Journal for Research in Mathematics Education, 28*(1), 26-47.

Malinski, M., Ross, A., Pannells, T., & McJunkin, M. (2006). Math anxiety in pre-service elementary school teachers. *Education, 127*(2), 274-279.

Mark, J., Goldenberg, E. P., Fries, M., Kang, J. M., & Cordner, T. (2014). *Transition to algebra*. Portsmouth, NH: Heinemann.

Mata, M. D. L., Monteiro, V., & Peixoto, F. (2012). Attitudes towards mathematics: Effects of individual, motivational, and social support factors. *Child Development Research,* Article ID 876028. doi: 10.1155/2012/876028

Matz-Costa, C., James, J. B., Ludlow, L. H., Brown, M., Besen, E., & Johnson, C. (2014). The meaning and measurement of productive engagement in later life. *Social Indicators Research, 118*(3), 1293-1314.

McLeod, D. B. (1992). "Role of affect in mathematics education: A reconceptualization," in *Handbook of Mathematics Teaching and Learning,* Ed. D. A. Grouws (New York, NY: NCTM), 575-596.

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*, 60-70.

Miller, H. & Bichsel, J. (2004). Anxiety, working memory, gender, and math performance. *Personality and Individual Differences, 37*, 591-606.

Morell, L., Collier, T., Black, P., & Wilson, M. (2017). A construct-modeling approach to develop a learning progression of how students understand the structure of matter. *Journal of Research in Science Teaching, 54*(8), 1024-1048.

Morgado, F. F. R., Meireles, J. F. F., Neves, C. M., Amaral, A. C. S., & Ferreira, M. E. C. (2017). Scale development: Ten main limitations and recommendations to improve future research practices. *Psicologia: Reflexão e Crítica, 30*(3), 1-20.

Morris, L. W., Davis, M. A., & Hutchings, C. H. (1981). Cognitive and emotional components of anxiety: Literature review and a revised worry-emotionality scale. *Journal of Educational Psychology, 73,* 541-555.

National Center for Educational Statistics (NCES). (2015, October 28). *The Nation's Report Card: 2015 Mathematics and Reading*. Retrieved from https://www.nationsreportcard.gov/reading_math_2015/#mathematics?grade=8

National Audit Office (2008). *Young People's Attitudes to Mathematics: A Research Study among 11-13 Year Olds on behalf of the National Audit House.* London: Author.

Núñez-Peña, M. I., & Suárez-Pellicioni, M. (2015). Processing of multi-digit additions in high math-anxious individuals: Psychophysiological evidence. *Frontiers in Psychology, 6*, 1268.

Nunnally, J. C. (1967). *Psychometric theory*. New York: McGraw Hill.

Organisation for Economic Co-operation and Development (OECD). (2013). *Ready to Learn: Students' Engagement, Drive and Self-Beliefs (Volume III),* PISA, OECD Publishing.

Osterlind, S. J. & Everson, H. T. (2009). *Differential item functioning*. Thousand Oaks, CA: Sage Publishing.

Pajares, F. & Miller, M. D. (1994). The role of self-efficacy and self-concept beliefs in mathematical problem solving: A path analysis. *Journal of Educational Psychology, 86*, 193-203.

Pane, J. F., Griffin, B. A., McCaffrey, D. F., & Karam, R. (2014). Effectiveness of Cognitive Tutor Algebra I at scale. *Educational Evaluation and Policy Analysis, 36*(2), 127–144.

Perry, A.B. (2004). Decreasing math anxiety in college students. *College Student Journal, 38*(2), 321-324.

Prieto, G. & Delgado, A. R. (2007). Measuring math anxiety (in Spanish) with the Rasch Rating Scale Model. *Journal of Applied Measurement, 8*, 149-160.

Rasch, G. (1960/1980). *Probabilistic models for some intelligence and attainment tests.* (Copenhagen, Danish Institute for Educational Research), expanded edition (1980) with foreword and afterword by B. D. Wright. Chicago, IL: The University of Chicago Press.

Rasch, G. (1966). An individualistic approach to item analysis. In P. F. Lazarsfeld & N. W. Henry (Eds.), *Readings in mathematical social science*, (pp. 84-108). Chicago, IL: Science Research Associates.

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

Richardson, F., & Woolfolk, C. (1980). Mathematics anxiety. In I. G. Sarason (Ed.), *Test anxiety: Theory, research, and applications* (pp. 271-288). Hillsdale, NJ: Erlbaum.

Sarason, I. G. (1986). Test anxiety, worry, and cognitive interference. In R. Schwarzer (Ed.), *Self-related cognitions in anxiety and motivation* (pp. 19-33). Hillsdale, NJ: Erlbaum.

Schmader, T. (2002). Gender identification moderates stereotype threat effects on women's math performance. *Journal of Experimental Social Psychology, 38*, 194-201.

Shen, T., Ludlow, L. H., Rogers, S., Millner, U., & Russinova, Z. (2018). The Community Readiness and Participation Scenario Scale: Unidimensional or multidimensional? Paper presented in a roundtable session at a meeting of the American Educational Research Association, New York, NY.

Sherman, J. & Fennema, E. (1978). Sex-related differences in mathematics achievement and related factors: A further study. *Journal for Research in Mathematics Education, 9*(3), 189-203.

Sondergeld, T. A. & Johnson, C. C. (2014). Using Rasch measurement for the development and use of affective assessments in science education research. *Science Education, 98*(4), 581-613.

Spelke, E. (2005). Sex differences in intrinsic aptitude for mathematics and science? A critical review. *American Psychologist, 60,* 950-958.

Spielberger, C. D. (1977). *The test anxiety inventory*. Palo Alto, CA: Consulting Psychologists Press.

Stankov, L. (2010). Unforgiving Confucian culture: A breeding ground for high academic achievement, test anxiety and self-doubt? *Learning and Individual Differences, 20*, 555–563.

Statistics Solutions. (n.d.). *Conduct and interpret a cluster analysis*. Retrieved February 5, 2021, from https://www.statisticssolutions.com/cluster-analysis-2/

Suinn, R., Taylor, S., & Edwards, R. (1998). Suinn Mathematics Anxiety Rating Scale for Elementary School Students (MARS-E): Psychometric and normative data. *Educational and Psychological Measurement, 48*, 979-986.

Taylor, J. A. (1953). A personality scale of manifest anxiety. *The Journal of Abnormal and Social Psychology, 48*(2), 285-290.

Tyron, G. (1980). The measurement and treatment of test anxiety. *Review of Educational Research, 50*, 353-372.

Vuilleumier, C. & Klein, K. (2015, April 16). Measuring mathematics testing confidence and anxiety: A scale analysis using Rasch modeling. Paper presented at the 2015 American Educational Research Association (AERA) Conference, Chicago, IL.

Walker, C. (2011). What's the DIF? Why differential item functioning analyses are an important part of instrument development and validation. *Journal of Psychoeducational Assessment, 29*, 364–376.

Wang, Z., Hart, S. A., Kovas, Y., Lukovski, S., Soden, B., Thompson, L. A., et al. (2014). Who is afraid of math? Two sources of genetic variance for mathematical anxiety. *Journal of Child Psychology and Psychiatry, 55*, 1056-1064.

Wigfield, A. & Meece, J. L. (1988). Math anxiety in elementary and secondary school students. *Educational Psychology, 80*, 210-216.

Williams, J. E. (1994). Anxiety measurement: Construct validity and test performance. *Measurement and Evaluation in Counseling and Development, 27*, 302-307.

Willis, G. B. (1999). *Cognitive interviewing: A "How To" guide*. Research Triangle Institute. 1999 Meeting of the American Statistical Association. Research Triangle Park, NC: Research Triangle Institute.

Wilson, M. (2005). *Constructing measures: An item response modeling approach*. Mahwah, NJ: Lawrence Erlbaum Associates.

Wine, J. (1971). Test anxiety and direction of attention. *Psychological Bulletin, 76*, 92-104.

Wolfe, E. W., & Smith, E. V. (2007). Instrument development tools and activities for measure validation using Rasch models: Part II—Validation activities. *Journal of Applied Measurement, 8*(2), 204-234.

Wright, B. D. (1980). "Foreward", and "Afterward". In Rasch, G. *Probabilistic Models for Some Intelligence and Attainment Tests*. Chicago, IL: The University of Chicago Press.

Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis*. Chicago, IL: MESA Press.

Wu, S. S., Barth, M., Amin, H., Malcarne, V., & Menon, V. (2012). Math anxiety in second and third graders and its relation to mathematical achievement. *Frontiers Developmental Psychology, 3*: 162.

Zohar, D. (1998). An additive model of test anxiety: Role of exam-specific expectations. *Journal of Educational Psychology, 90*(2), 330-340.

APPENDIX A

Figure A1

Comparative Math Anxiety (Emotional-Attitudinal) Dendrogram, R2

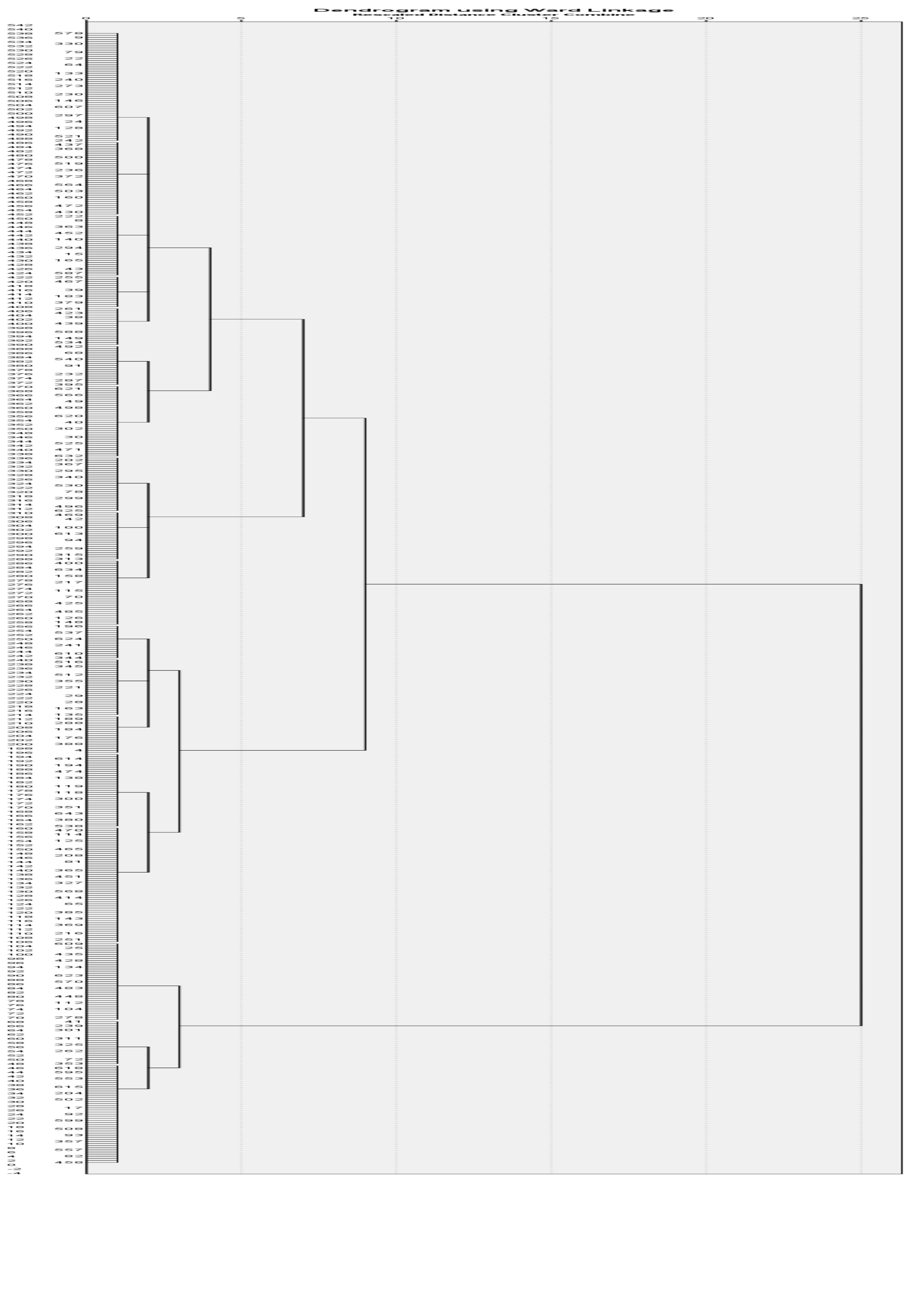


Figure A2

Comparative Math Anxiety (Mental-Cognitive) Dendrogram, R2

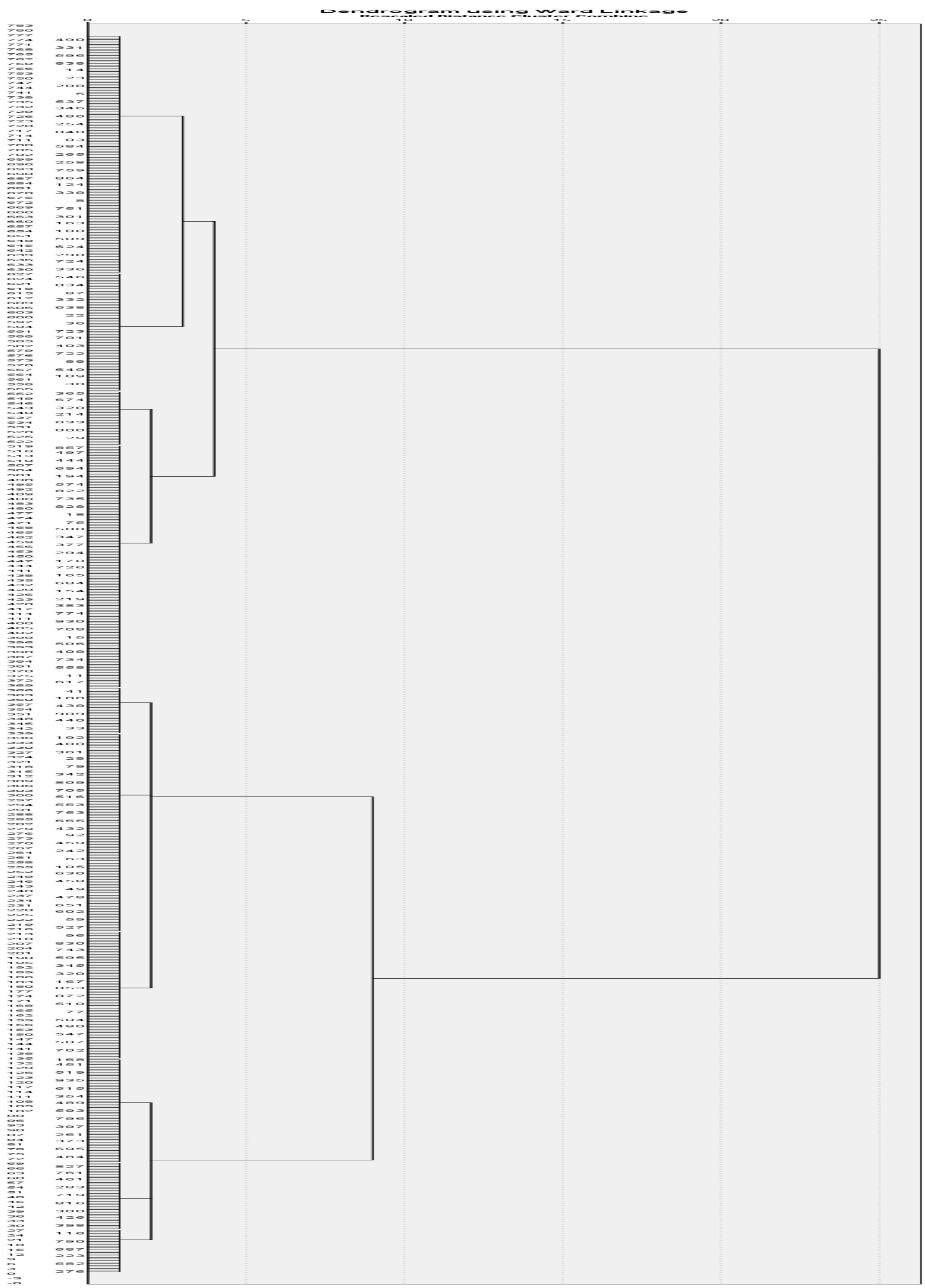


Figure A3

Comparative Math Anxiety (Physical-Somatic) Dendrogram, R2

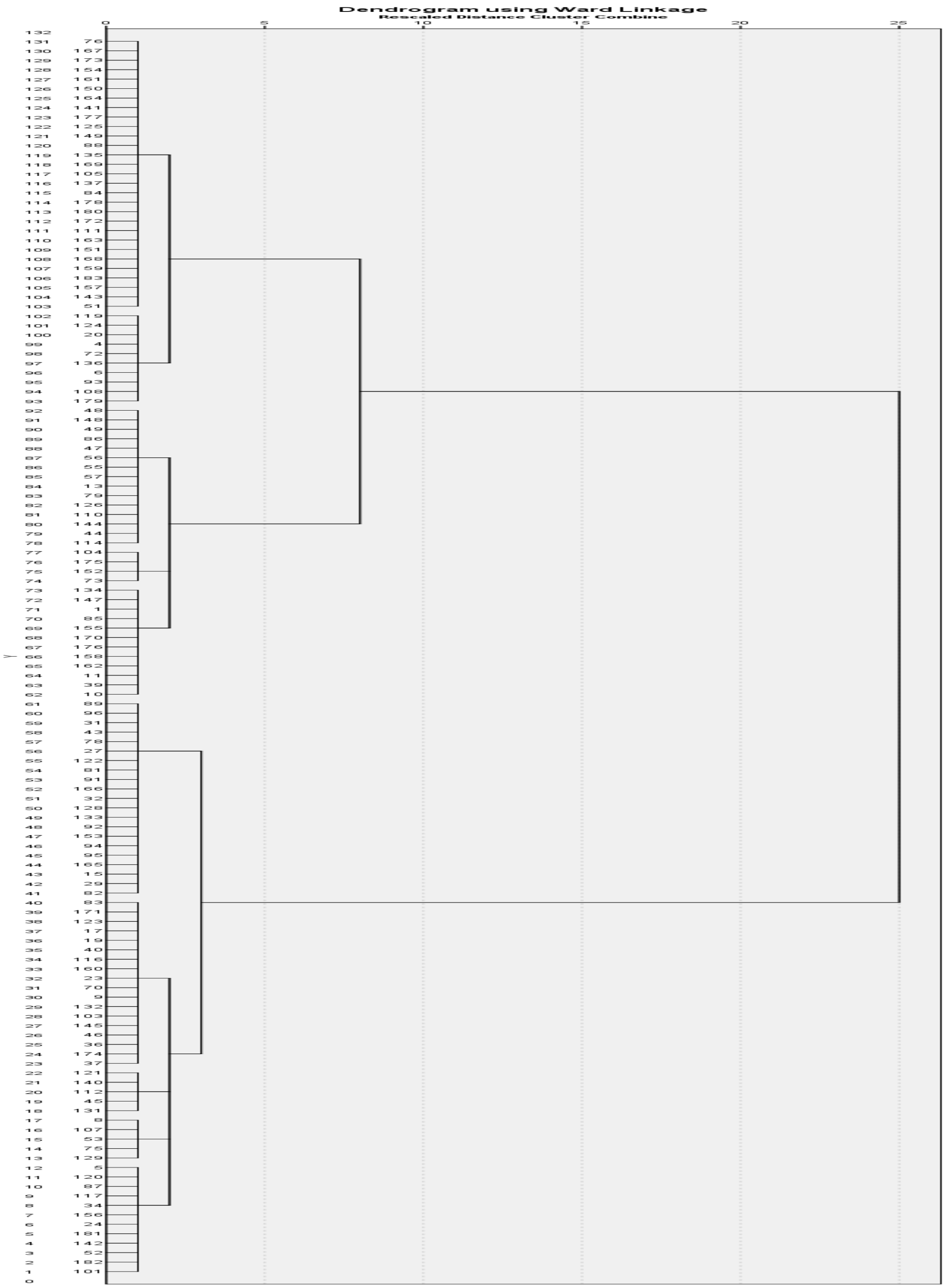


Table A1

Comparative Math Anxiety (Mental-Cognitive) Person Fit Statistics, R2 (Expected)

Person	Measure	INFIT		OUTFIT	
		MNSQ	*ZSTD*	*MNSQ*	*ZSTD*
198	-2.07	3.56	2.9	3.10	2.5
306	-1.32	3.48	3.8	3.29	3.6
322	-1.32	3.48	3.8	3.29	3.6
313	-0.08	3.04	3.4	3.04	3.4
220	0.02	3.01	3.4	3.02	3.4
63	-0.38	2.9	3.3	2.90	3.3
229	-0.59	2.71	3.1	2.74	3.1
28	-0.08	2.70	3	2.71	3.0
286	-0.59	2.67	3	2.70	3.1
10	0.32	2.68	3.1	2.69	3.1
324	-0.18	2.67	3	2.69	3.0
184	-0.38	2.58	2.9	2.59	2.9
62	0.22	2.55	2.9	2.56	2.9
1	-0.28	2.54	2.8	2.54	2.8
9	0.12	2.41	2.6	2.41	2.6
164	0.02	2.35	2.6	2.35	2.6
338	-0.28	2.35	2.5	2.35	2.6
38	-1.47	2.33	2.3	2.07	2.0
156	1.06	2.32	2.6	2.29	2.6
238	-0.08	2.3	2.5	2.29	2.5
210	0.02	2.29	2.5	2.29	2.5
214	0.94	2.29	2.6	2.25	2.5
188	-0.08	2.24	2.4	2.24	2.4
162	-0.38	2.2	2.3	2.21	2.4
65	0.22	2.19	2.4	2.19	2.4
333	-0.08	2.19	2.3	2.19	2.3
58	-0.18	2.17	2.3	2.18	2.3
185	-0.70	2.16	2.3	2.17	2.3
61	-0.28	2.14	2.3	2.14	2.3
17	-1.05	2.11	2.2	2.07	2.2
212	-0.18	2.10	2.2	2.10	2.2
148	-0.08	2.09	2.2	2.08	2.2
355	0.12	2.01	2.1	2.01	2.1
312	0.22	1.99	2.1	1.99	2.1
194	-0.48	1.97	2	1.97	2.0
329	-0.28	1.89	1.9	1.89	1.9
136	-0.18	1.88	1.9	1.88	1.8
227	0.32	1.84	1.8	1.85	1.8
275	-0.18	1.84	1.8	1.84	1.8
337	-0.18	1.82	1.8	1.82	1.8

101	-0.70	1.81	1.8	1.77	1.7
102	-0.81	1.81	1.8	1.81	1.8
358	0.32	1.80	1.8	1.79	1.8
52	-0.18	1.79	1.7	1.78	1.7
104	0.12	1.77	1.7	1.78	1.7
277	0.12	1.77	1.7	1.77	1.7
315	0.42	1.77	1.7	1.77	1.7
50	-0.48	1.75	1.7	1.74	1.6
265	0.12	1.75	1.7	1.75	1.7
30	-0.18	1.74	1.6	1.73	1.6
276	0.32	1.74	1.7	1.72	1.6
23	0.12	1.71	1.6	1.72	1.6
241	-2.76	1.72	1	1.47	0.8
336	-0.28	1.72	1.6	1.72	1.6
280	0.12	1.69	1.6	1.70	1.6
363	-0.28	1.67	1.5	1.68	1.5
32	-0.08	1.66	1.5	1.66	1.5
13	0.22	1.65	1.5	1.65	1.5
294	-0.28	1.64	1.5	1.65	1.5
78	-0.28	1.59	1.4	1.61	1.4
197	0.22	1.61	1.4	1.61	1.4
248	-0.08	1.61	1.4	1.61	1.4
60	-0.48	1.59	1.4	1.59	1.4
170	-0.59	1.57	1.3	1.59	1.4
147	-0.92	1.56	1.3	1.48	1.2
328	-0.28	1.54	1.3	1.54	1.3
33	-0.28	1.51	1.2	1.53	1.3
151	-0.18	1.50	1.2	1.50	1.2
54	0.12	1.49	1.2	1.49	1.2
107	0.52	1.49	1.2	1.48	1.2
288	-0.38	1.48	1.2	1.47	1.1
193	-0.28	1.45	1.1	1.45	1.1
215	-0.18	1.44	1.1	1.45	1.1
155	0.42	1.44	1.1	1.43	1.1
149	-2.36	1.41	0.8	1.43	0.8
70	-0.70	1.42	1.1	1.41	1.1
130	0.32	1.39	1	1.41	1.0
157	-0.81	1.39	1	1.41	1.0
352	-1.18	1.38	1	1.40	1.0
128	-1.64	1.39	0.9	1.34	0.8
154	-0.08	1.38	1	1.38	1.0
270	-0.81	1.38	1	1.36	1.0
41	0.42	1.37	1	1.36	1.0
117	-1.18	1.36	0.9	1.37	0.9
204	-0.38	1.37	1	1.35	0.9
219	0.22	1.37	1	1.37	1.0

320	-0.28	1.37	0.9	1.35	0.9
359	0.52	1.37	1	1.35	1.0
301	-0.18	1.36	0.9	1.36	0.9
331	0.22	1.36	0.9	1.35	0.9
85	0.12	1.35	0.9	1.35	0.9
299	-0.48	1.35	0.9	1.34	0.9
316	-0.81	1.34	0.9	1.35	0.9
272	0.52	1.34	0.9	1.34	0.9
89	0.22	1.32	0.9	1.33	0.9
273	0.22	1.32	0.9	1.33	0.9
77	0.22	1.32	0.9	1.32	0.9
309	-0.59	1.29	0.8	1.32	0.9
356	-0.18	1.31	0.8	1.32	0.8
191	-0.28	1.31	0.8	1.31	0.8
262	0.42	1.31	0.9	1.31	0.8
203	0.02	1.30	0.8	1.30	0.8
244	0.02	1.29	0.8	1.29	0.8
228	-0.28	1.28	0.8	1.27	0.7
105	-2.07	1.20	0.5	1.27	0.6
27	0.42	1.22	0.7	1.26	0.7
327	-1.84	1.26	0.6	1.19	0.5
189	0.62	1.25	0.7	1.24	0.7
290	0.02	1.25	0.7	1.25	0.7
82	-0.18	1.23	0.7	1.23	0.7
59	-0.38	1.21	0.6	1.20	0.6
140	-1.84	1.21	0.6	1.15	0.4
318	-0.81	1.21	0.6	1.21	0.6

APPENDIX B

Table B1

Comparative Math Anxiety (Emotional-Attitudinal) Residual Analysis, Total Variance Explained, R3

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	2.636	32.956	32.956	2.636	32.956	32.956
2	1.260	15.753	48.709	1.260	15.753	48.709
3	1.247	15.586	64.295	1.247	15.586	64.295
4	.904	11.299	75.594			
5	.759	9.487	85.082			
6	.670	8.369	93.451			
7	.495	6.193	99.644			
8	.028	.356	100			

Figure B1

Comparative Math Anxiety (Emotional-Attitudinal) Residual Analysis, Scree Plot, R3

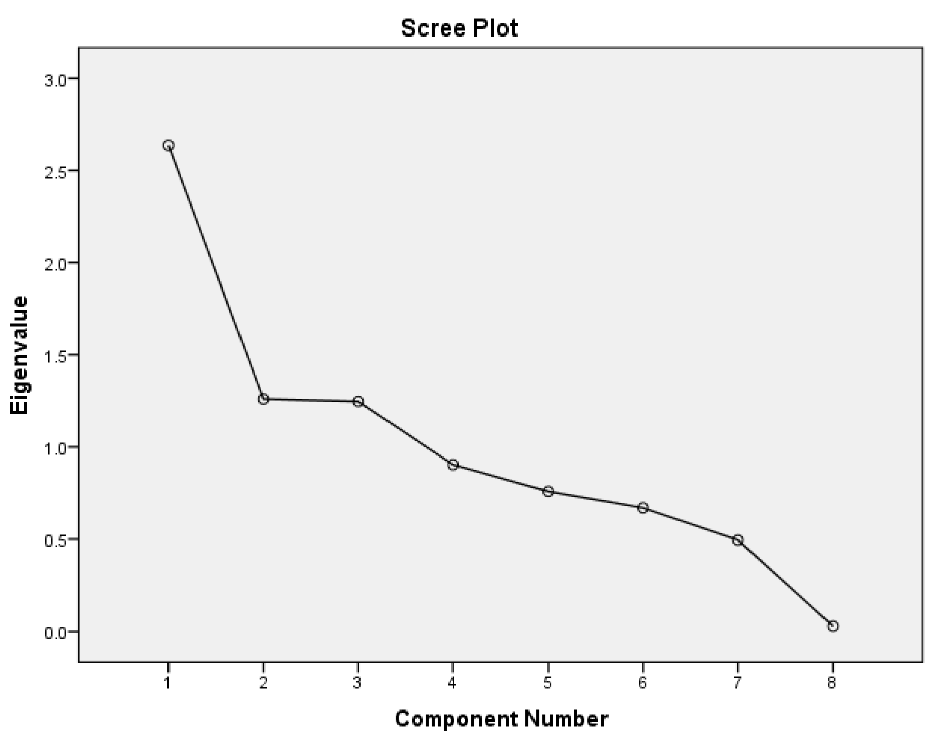


Table B2

Comparative Math Anxiety (Emotional-Attitudinal) Residual Analysis, Component Matrix, R3

Item	Component Loading
A1_3_Viktoria	-.089
A1_4_Daniel	-.822
A1_5_Ruby	.634
A1_7_David	-.294
A1_10_Andrea	.463
A1_11_Chloe	-.762
A1_12_Gabriel	.436
A1_13_Larissa	.692

Figure B2

Comparative Math Anxiety (Emotional-Attitudinal) Residual Analysis, Component Plot, R3

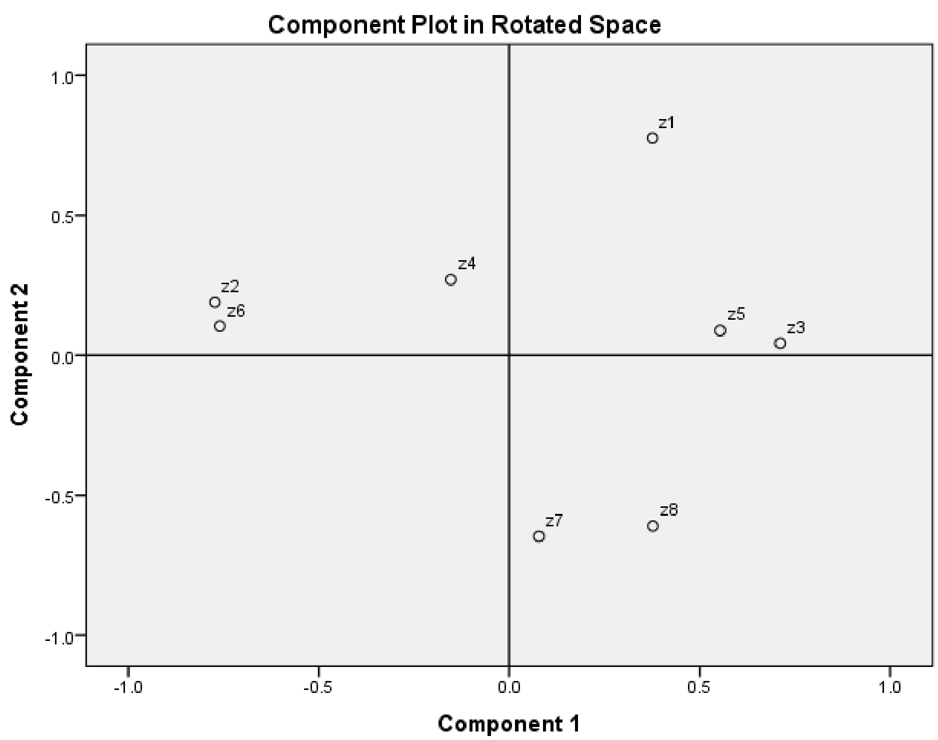


Table B3

Comparative Math Anxiety (Mental-Cognitive) Residual Analysis, Total Variance Explained, R3

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	2.204	31.486	31.486	2.204	31.486	31.486
2	1.255	17.932	49.419	1.255	17.932	49.419
3	1.161	16.579	65.997	1.161	16.579	65.997
4	1.051	15.021	81.018	1.051	15.021	81.018
5	.787	11.242	92.260			
6	.524	7.481	99.742			
7	.018	.258	100			

Figure B3

Comparative Math Anxiety (Mental-Cognitive) Residual Analysis, Scree Plot, R3

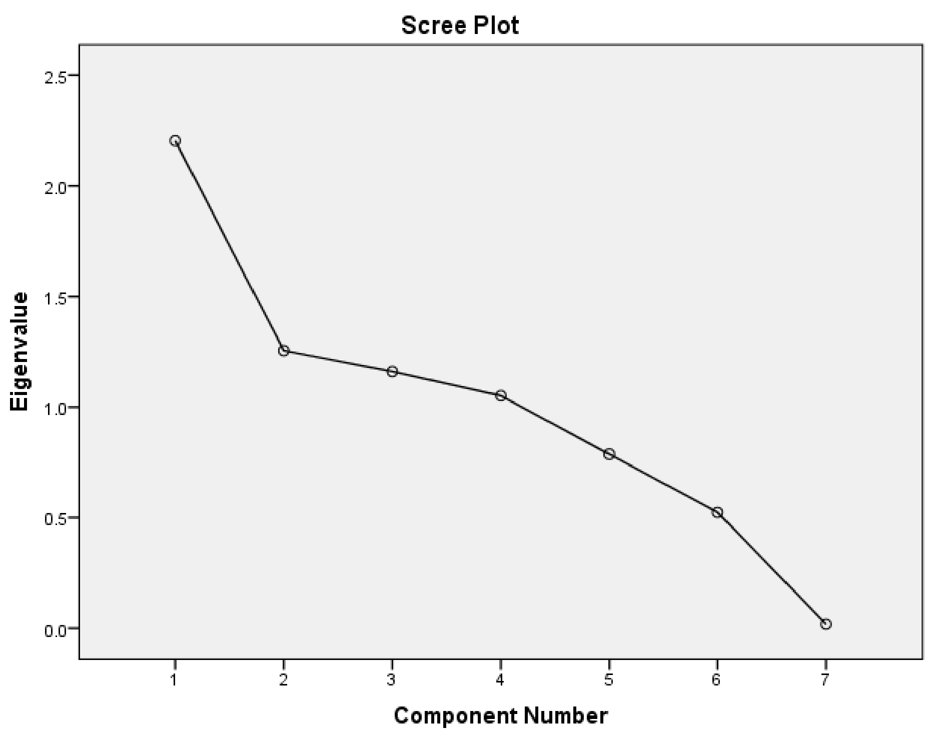


Table B4

Comparative Math Anxiety (Mental-Cognitive) Residual Analysis, Component Matrix, R3

Item	Component Loading
A2_2_Ana	.385
A2_3_Carlos	-.767
A2_5_Petra	.013
A2_7_Lena	-.815
A2_8_Leon	.481
A2_9_Emilie	.440
A2_10_Joseph	.615

Figure B4

Comparative Math Anxiety (Mental-Cognitive) Residual Analysis, Component Plot, R3

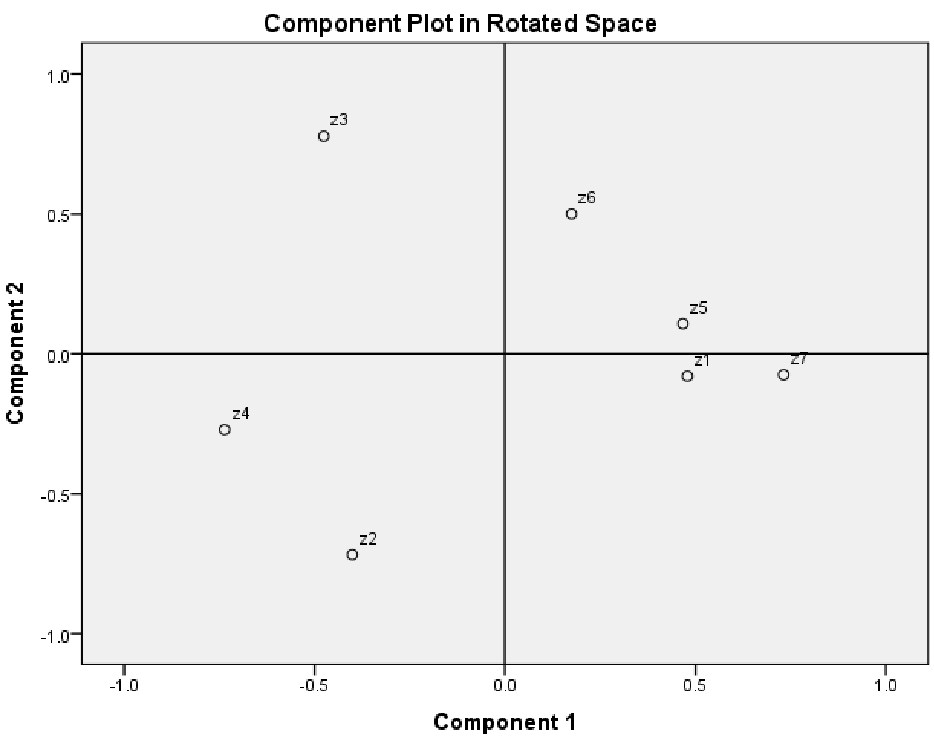


Table B5

Comparative Math Anxiety (Physical-Somatic) Residual Analysis, Total Variance Explained, R3

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	2.848	35.599	35.599	2.848	35.599	35.599
2	1.141	14.258	49.857	1.141	14.258	49.857
3	1.049	13.118	62.974	1.049	13.118	62.974
4	.964	12.045	75.020			
5	.865	10.810	85.830			
6	.692	8.648	94.478			
7	.379	4.733	99.211			
8	.063	.789	100.000			

Figure B5

Comparative Math Anxiety (Physical-Somatic) Residual Analysis, Scree Plot, R3

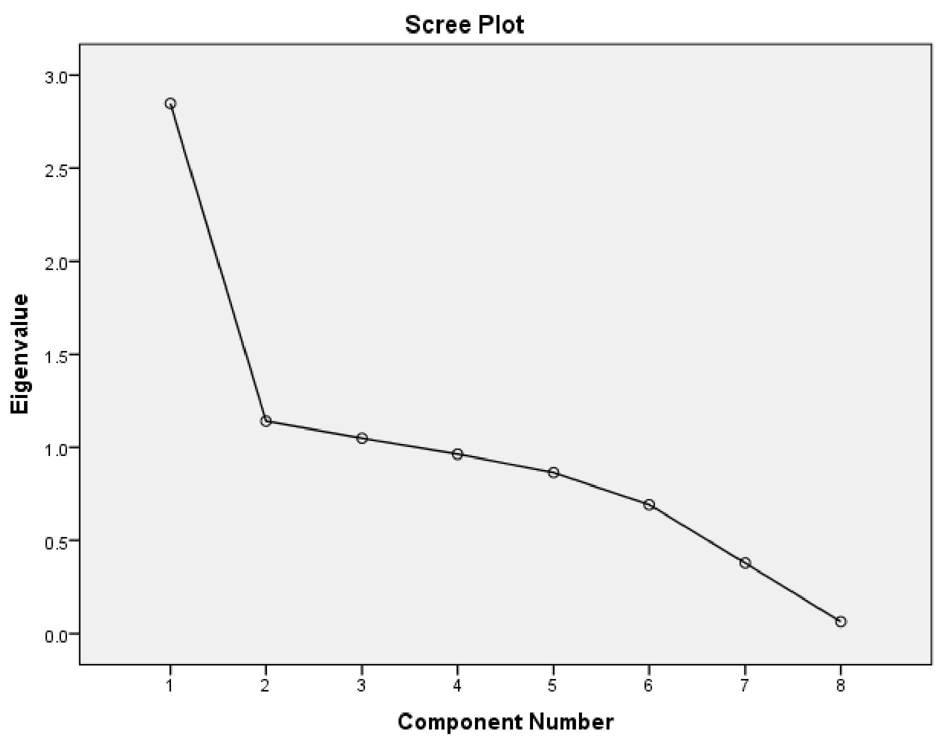


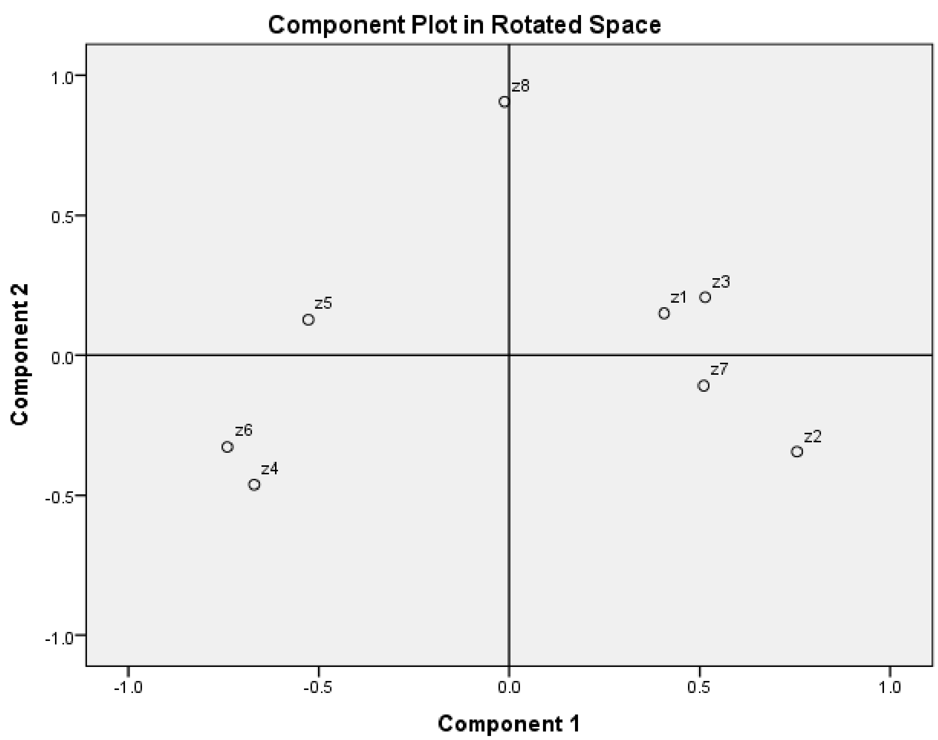
Table B6

Comparative Math Anxiety (Physical-Somatic) Residual Analysis, Component Matrix, R3

Item	Component Loading
A3_1_Ivan	.455
A3_2_Gabrielle	.554
A3_3_Laila	.583
A3_4_Jay	-.849
A3_7_Nora	-.439
A3_8_Romita	-.852
A3_9_Noah	.432
A3_10_Erica	.411

Figure B6

Comparative Math Anxiety (Physical-Somatic) Residual Analysis, Component Plot, R3



APPENDIX C

Table C1

Comparative Math Anxiety (Emotional-Attitudinal) DIF Analysis, Grade

Item		Person Class	Observations		Baseline		DIF					
#	Name		Count	Avg.	Expect	Measure	Score	Measure	Size	S.E.	t	Prob.
1	A1_3	9	28	1.39	1.48	0.16	-0.09	0.34	0.18	0.27	0.66	0.5129
		10	16	1.56	1.47	0.16	0.09	-0.04	-0.2	0.37	-0.55	0.5919
		11	51	1.63	1.63	0.16	0	0.16	0	0.2	0	1
		12	27	1.37	1.34	0.16	0.03	0.1	-0.06	0.28	-0.23	0.8226
2	A1_4	9	27	2.11	2.28	-1.46	-0.17	-1.12	0.34	0.27	1.23	0.2296
		10	16	2.63	2.23	-1.46	0.39	-2.29	-0.82	0.37	-2.25	0.0414
		11	53	2.38	2.43	-1.46	-0.05	-1.36	0.1	0.2	0.51	0.6144
		12	27	2.19	2.15	-1.46	0.03	-1.53	-0.06	0.27	-0.24	0.8151
3	A1_5	9	28	0.71	0.68	1.97	0.03	1.88	-0.08	0.31	-0.27	0.7896
		10	16	0.69	0.73	1.97	-0.05	2.1	0.14	0.43	0.31	0.758
		11	52	0.9	0.82	1.97	0.09	1.75	-0.22	0.22	-1	0.3242
		12	27	0.44	0.62	1.97	-0.18	2.57	0.6	0.37	1.61	0.1198
4	A1_7	9	28	1.68	1.63	-0.13	0.05	-0.24	-0.11	0.27	-0.39	0.6966
		10	16	1.5	1.61	-0.13	-0.11	0.1	0.23	0.37	0.62	0.5437
		11	51	1.78	1.78	-0.13	0	-0.13	0	0.2	0	1
		12	27	1.48	1.48	-0.13	0	-0.13	0	0.28	0	1
5	A1_10	9	28	1.21	1.22	0.7	-0.01	0.7	0	0.28	0	1
		10	16	1.25	1.23	0.7	0.02	0.66	-0.04	0.38	-0.11	0.9113
		11	52	1.37	1.36	0.7	0	0.7	0	0.2	0	1
		12	26	1.12	1.13	0.7	-0.01	0.74	0.03	0.3	0.11	0.9114
6	A1_11	9	28	2.25	2.32	-1.54	-0.07	-1.4	0.14	0.27	0.52	0.6089
		10	16	2.56	2.27	-1.54	0.29	-2.15	-0.61	0.36	-1.68	0.1144
		11	52	2.37	2.47	-1.54	-0.11	-1.32	0.22	0.2	1.09	0.2791
		12	27	2.3	2.19	-1.54	0.11	-1.75	-0.21	0.27	-0.77	0.4484
7	A1_12	9	28	1.79	1.67	-0.22	0.12	-0.46	-0.23	0.27	-0.87	0.3913
		10	16	1.38	1.65	-0.22	-0.27	0.38	0.6	0.37	1.6	0.133
		11	52	1.87	1.81	-0.22	0.05	-0.32	-0.1	0.2	-0.53	0.6005
		12	26	1.42	1.49	-0.22	-0.06	-0.09	0.13	0.28	0.46	0.6495
8	A1_13	9	28	1.43	1.31	0.52	0.12	0.27	-0.24	0.27	-0.9	0.3745
		10	16	0.94	1.31	0.52	-0.38	1.42	0.9	0.4	2.25	0.0408
		11	52	1.46	1.45	0.52	0.01	0.52	0	0.2	0	1
		12	27	1.26	1.18	0.52	0.08	0.34	-0.18	0.28	-0.62	0.5431

Table C2

Comparative Math Anxiety (Emotional-Attitudinal) DIF Analysis, Gender

Item		Person	Obs-Exp	DIF		Person	Obs-Exp	DIF		DIF	Joint	Rasch-Welch			Mantel	
#	Name	Class	Avg.	Measure	S.E.	Class	Avg.	Measure	S.E.	Contrast	S.E.	t	df	Prob.	Chi-sq	Prob.
1	A1_3	0	-0.05	0.25	0.2	1	0.04	0.08	0.18	0.17	0.26	0.65	111	0.5144	1.3942	0.2377
2	A1_4	0	-0.09	-1.25	0.19	1	0.07	-1.58	0.18	0.33	0.26	1.25	114	0.2156	2.4417	0.1181
3	A1_5	0	-0.05	2.07	0.24	1	0.03	1.85	0.2	0.22	0.31	0.73	110	0.4665	0.0155	0.901
4	A1_7	0	-0.02	-0.08	0.2	1	0.02	-0.17	0.17	0.08	0.26	0.32	111	0.7507	1.6954	0.1929
5	A1_10	0	-0.01	0.68	0.2	1	0.01	0.68	0.18	0	0.27	0	113	1	0.1347	0.7136
6	A1_11	0	0.12	-1.76	0.19	1	-0.09	-1.33	0.17	-0.43	0.26	-1.64	113	0.1029	0.3729	0.5414
7	A1_12	0	-0.02	-0.19	0.19	1	0.01	-0.24	0.18	0.05	0.26	0.21	113	0.8373	0.0086	0.9263
8	A1_13	0	0.1	0.31	0.2	1	-0.09	0.71	0.18	-0.41	0.27	-1.52	114	0.1325	4.5897	0.0322

Table C3

Comparative Math Anxiety (Emotional-Attitudinal) DIF Analysis, English-Language Proficient

Item		Person	Obs-Exp	DIF		Person	Obs-Exp	DIF		DIF	Joint	Rasch-Welch			Mantel	
#	Name	Class	Avg.	Measure	S.E.	Class	Avg.	Measure	S.E.	Contrast	S.E.	t	df	Prob.	Chi-sq	Prob.
1	A1_3	0	0	0.16	0.14	1	0.01	0.16	0.41	0	0.43	0	13	1	0.0271	0.8693
2	A1_4	0	0.07	-1.6	0.13	1	-0.67	-0.14	0.43	-1.45	0.45	-3.26	12	0.0068	3.4743	0.0623
3	A1_5	0	-0.06	2.15	0.16	1	0.58	0.67	0.42	1.49	0.45	3.31	14	0.0051	2.8075	0.0938
4	A1_7	0	0.02	-0.18	0.14	1	-0.22	0.32	0.41	-0.5	0.43	-1.15	13	0.2693	1.9096	0.167
5	A1_10	0	-0.07	0.85	0.14	1	0.6	-0.51	0.41	1.37	0.43	3.18	13	0.0073	10.3183	0.0013
6	A1_11	0	0.08	-1.71	0.14	1	-0.76	-0.02	0.41	-1.69	0.43	-3.92	13	0.0017	4.8712	0.0273
7	A1_12	0	-0.01	-0.22	0.14	1	0.06	-0.35	0.41	0.13	0.43	0.3	13	0.7703	0.4593	0.4979
8	A1_13	0	-0.04	0.6	0.14	1	0.34	-0.18	0.41	0.78	0.43	1.82	13	0.0925	0.822	0.3646

Table C4

Comparative Math Anxiety (Mental-Cognitive) DIF Analysis, Grade

Item		Person Class	Observations		Baseline		DIF					
#	Name		Count	Avg.	Expect	Measure	Score	Measure	Size	S.E.	t	Prob.
1	A2_2	9	29	1.66	1.68	0.17	-0.03	0.24	0.07	0.28	0.24	0.8131
		10	16	1.31	1.41	0.17	-0.1	0.42	0.26	0.41	0.63	0.5397
		11	53	1.58	1.68	0.17	-0.09	0.38	0.21	0.21	1.03	0.3088
		12	27	1.59	1.32	0.17	0.27	-0.47	-0.63	0.3	-2.14	0.0421
2	A2_3	9	29	2.31	2.35	-1.39	-0.04	-1.3	0.09	0.29	0.3	0.7653
		10	16	2.44	2.06	-1.39	0.37	-2.27	-0.88	0.39	-2.25	0.041
		11	53	2.28	2.35	-1.39	-0.06	-1.24	0.15	0.21	0.71	0.4796
		12	27	1.93	1.98	-1.39	-0.05	-1.26	0.13	0.3	0.44	0.6628
3	A2_5	9	29	1.76	1.85	-0.22	-0.09	-0.01	0.22	0.28	0.76	0.453
		10	16	1.63	1.57	-0.22	0.06	-0.36	-0.14	0.39	-0.36	0.7244
		11	52	1.88	1.85	-0.22	0.04	-0.31	-0.09	0.21	-0.41	0.6837
		12	27	1.48	1.49	-0.22	-0.01	-0.22	0	0.3	0	1
4	A2_7	9	29	2.41	2.28	-1.22	0.14	-1.55	-0.33	0.29	-1.15	0.2604
		10	16	2.19	1.99	-1.22	0.2	-1.68	-0.45	0.38	-1.19	0.2535
		11	53	2.17	2.27	-1.22	-0.1	-0.98	0.24	0.21	1.17	0.2493
		12	27	1.85	1.91	-1.22	-0.06	-1.08	0.14	0.3	0.47	0.6441
5	A2_8	9	29	1.76	1.8	-0.11	-0.05	-0.01	0.11	0.28	0.38	0.7083
		10	16	1.25	1.52	-0.11	-0.27	0.59	0.7	0.41	1.72	0.1083
		11	53	1.96	1.8	-0.11	0.16	-0.49	-0.38	0.21	-1.81	0.0769
		12	27	1.33	1.44	-0.11	-0.11	0.15	0.26	0.3	0.88	0.3879
6	A2_9	9	29	0.86	0.77	2.53	0.09	2.25	-0.28	0.32	-0.89	0.3804
		10	16	0.56	0.63	2.53	-0.06	2.78	0.25	0.5	0.5	0.6259
		11	53	0.7	0.71	2.53	-0.01	2.57	0.03	0.23	0.15	0.8839
		12	27	0.41	0.45	2.53	-0.05	2.71	0.18	0.38	0.47	0.6435
7	A2_10	9	29	1.62	1.65	0.24	-0.03	0.32	0.07	0.29	0.26	0.7979
		10	16	1.19	1.38	0.24	-0.19	0.76	0.52	0.42	1.24	0.2338
		11	53	1.72	1.65	0.24	0.07	0.08	-0.17	0.21	-0.8	0.4295
		12	27	1.3	1.29	0.24	0	0.24	0	0.3	0	1

Table C5

Comparative Math Anxiety (Mental-Cognitive) DIF Analysis, Gender

Item		Person Class	Obs-Exp Avg.	DIF		Person Class	Obs-Exp Avg.	DIF		DIF Contrast	Joint S.E.	Rasch-Welch			Mantel	
#	Name			Measure	S.E.			Measure	S.E.			t	df	Prob.	Chi-sq	Prob.
1	A2_2	0	0.04	0.07	0.21	1	-0.03	0.24	0.18	-0.17	0.28	-0.6	114	0.5489	0.712	0.3988
2	A2_3	0	0.06	-1.53	0.21	1	-0.05	-1.26	0.19	-0.27	0.28	-0.97	115	0.3316	0.583	0.4452
3	A2_5	0	-0.08	-0.05	0.21	1	0.06	-0.37	0.18	0.32	0.28	1.16	112	0.2481	0.7526	0.3856
4	A2_7	0	0.05	-1.31	0.21	1	-0.04	-1.12	0.18	-0.19	0.28	-0.69	115	0.4896	0.0369	0.8477
5	A2_8	0	0.01	-0.15	0.21	1	-0.01	-0.1	0.18	-0.05	0.28	-0.16	114	0.8711	0.0097	0.9216
6	A2_9	0	-0.07	2.75	0.26	1	0.05	2.37	0.21	0.38	0.33	1.15	109	0.2526	1.224	0.2686
7	A2_10	0	-0.02	0.29	0.21	1	0.02	0.21	0.18	0.09	0.28	0.3	114	0.7611	0.0022	0.9623

Table C6

Comparative Math Anxiety (Mental-Cognitive) DIF Analysis, English-Language Proficient

Item		Person Class	Obs-Exp Avg.	DIF		Person Class	Obs-Exp Avg.	DIF		DIF Contrast	Joint S.E.	Rasch-Welch			Mantel	
#	Name			Measure	S.E.			Measure	S.E.			t	df	Prob.	Chi-sq	Prob.
1	A2_2	0	-0.03	0.24	0.14	1	0.27	-0.48	0.44	0.71	0.47	1.53	13	0.1491	1.7912	0.1808
2	A2_3	0	0.03	-1.47	0.14	1	-0.31	-0.67	0.44	-0.8	0.46	-1.72	13	0.109	0.5426	0.4614
3	A2_5	0	0.01	-0.26	0.14	1	-0.14	0.12	0.45	-0.37	0.47	-0.8	13	0.4392	0.1036	0.7476
4	A2_7	0	0.04	-1.32	0.14	1	-0.4	-0.28	0.44	-1.04	0.47	-2.24	13	0.0433	2.6343	0.1046
5	A2_8	0	-0.02	-0.07	0.14	1	0.15	-0.48	0.44	0.4	0.47	0.86	13	0.4041	0.0545	0.8154
6	A2_9	0	-0.02	2.58	0.17	1	0.13	2.18	0.47	0.4	0.5	0.79	13	0.4426	0.0005	0.983
7	A2_10	0	-0.03	0.32	0.14	1	0.3	-0.48	0.44	0.8	0.47	1.71	13	0.1104	3.9135	0.0479

Table C7

Comparative Math Anxiety (Physical-Somatic) DIF Analysis, Grade

Item		Person Class	Observations		Baseline		DIF					
#	Name		Count	Avg.	Expect	Measure	Score	Measure	Size	S.E.	t	Prob.
1	A3_1	9	29	1	0.89	0.96	0.11	0.73	-0.24	0.27	-0.87	0.3935
		10	16	0.88	0.89	0.96	-0.02	1.01	0.05	0.39	0.12	0.9076
		11	53	0.98	0.96	0.96	0.02	0.92	-0.04	0.21	-0.22	0.8304
		12	27	0.63	0.78	0.96	-0.15	1.37	0.4	0.33	1.23	0.2291
2	A3_2	9	29	1	0.86	1.04	0.14	0.73	-0.31	0.27	-1.14	0.2633
		10	16	0.63	0.86	1.04	-0.24	1.69	0.65	0.43	1.49	0.1585
		11	53	0.94	0.93	1.04	0.02	1	-0.03	0.21	-0.17	0.8663
		12	27	0.7	0.75	1.04	-0.05	1.16	0.12	0.32	0.38	0.7076
3	A3_3	9	28	0.39	0.44	2.08	-0.04	2.23	0.15	0.36	0.41	0.6842
		10	16	0.38	0.5	2.08	-0.13	2.58	0.49	0.52	0.95	0.3589
		11	53	0.51	0.52	2.08	-0.02	2.13	0.05	0.24	0.2	0.8457
		12	27	0.56	0.41	2.08	0.14	1.59	-0.49	0.34	-1.45	0.1608
4	A3_4	9	29	2.41	2.42	-2.16	-0.01	-2.16	0	0.27	0	1
		10	16	2.63	2.39	-2.16	0.24	-2.68	-0.52	0.38	-1.39	0.1869
		11	53	2.43	2.48	-2.16	-0.05	-2.06	0.1	0.2	0.49	0.6258
		12	27	2.22	2.26	-2.16	-0.04	-2.07	0.08	0.28	0.3	0.7669
5	A3_7	9	29	1.83	1.78	-0.86	0.05	-0.95	-0.09	0.26	-0.34	0.7329
		10	16	1.88	1.75	-0.86	0.12	-1.1	-0.24	0.35	-0.68	0.5079
		11	53	1.79	1.85	-0.86	-0.06	-0.75	0.11	0.19	0.57	0.5701
		12	26	1.65	1.67	-0.86	-0.01	-0.83	0.03	0.27	0.09	0.9269
6	A3_8	9	29	2.28	2.4	-2.11	-0.12	-1.85	0.25	0.27	0.95	0.3487
		10	16	2.75	2.36	-2.11	0.39	-2.97	-0.86	0.38	-2.24	0.0415
		11	52	2.38	2.46	-2.11	-0.08	-1.94	0.17	0.2	0.84	0.4037
		12	27	2.3	2.24	-2.11	0.06	-2.23	-0.12	0.28	-0.44	0.6667
7	A3_9	9	29	0.93	1	0.73	-0.07	0.88	0.15	0.28	0.55	0.5877
		10	16	1	0.99	0.73	0.01	0.73	0	0.38	0	1
		11	53	1.11	1.07	0.73	0.04	0.63	-0.09	0.2	-0.47	0.6406
		12	27	0.85	0.87	0.73	-0.02	0.78	0.05	0.3	0.17	0.8659
8	A3_10	9	29	1.14	1.2	0.31	-0.06	0.43	0.12	0.27	0.45	0.656
		10	16	0.81	1.18	0.31	-0.37	1.17	0.86	0.4	2.14	0.0503
		11	53	1.38	1.27	0.31	0.11	0.08	-0.23	0.2	-1.16	0.2512
		12	27	1.11	1.05	0.31	0.06	0.18	-0.13	0.28	-0.46	0.6524

Table C8

Comparative Math Anxiety (Physical-Somatic) DIF Analysis, Gender

Item		Person	Obs-Exp	DIF		Person	Obs-Exp	DIF		DIF	Joint	Rasch-Welch			Mantel	
#	Name	Class	Avg.	Measure	S.E.	Class	Avg.	Measure	S.E.	Contrast	S.E.	t	df	Prob.	Chi-sq	Prob.
1	A3_1	0	0.23	0.43	0.2	1	-0.19	1.41	0.19	-0.97	0.27	-3.54	119	0.0006	12.8243	0.0003
2	A3_2	0	0.05	0.92	0.21	1	-0.04	1.12	0.19	-0.2	0.28	-0.72	115	0.4743	0.7478	0.3872
3	A3_3	0	-0.02	2.17	0.27	1	0.01	2.06	0.21	0.11	0.34	0.32	106	0.7489	0.1049	0.746
4	A3_4	0	-0.06	-2.04	0.19	1	0.05	-2.26	0.18	0.22	0.26	0.85	116	0.3969	0.605	0.4367
5	A3_7	0	-0.05	-0.75	0.19	1	0.04	-0.93	0.17	0.18	0.25	0.71	116	0.4777	3.3279	0.0681
6	A3_8	0	-0.11	-1.85	0.19	1	0.09	-2.26	0.18	0.41	0.26	1.56	115	0.1212	1.6519	0.1987
7	A3_9	0	0.07	0.55	0.2	1	-0.06	0.82	0.18	-0.27	0.27	-0.99	116	0.3259	0.7806	0.377
8	A3_10	0	-0.11	0.55	0.2	1	0.09	0.12	0.17	0.43	0.26	1.62	114	0.1074	1.7135	0.1905

Table C9

Comparative Math Anxiety (Physical-Somatic) DIF Analysis, English-Language Proficient

Item		Person	Obs-Exp	DIF		Person	Obs-Exp	DIF		DIF	Joint	Rasch-Welch			Mantel	
#	Name	Class	Avg.	Measure	S.E.	Class	Avg.	Measure	S.E.	Contrast	S.E.	t	df	Prob.	Chi-sq	Prob.
1	A3_1	0	-0.04	1.07	0.15	1	0.41	0.1	0.41	0.97	0.44	2.23	13	0.0442	2.1856	0.1393
2	A3_2	0	-0.04	1.14	0.15	1	0.36	0.27	0.41	0.86	0.44	1.97	13	0.0702	5.6697	0.0173
3	A3_3	0	-0.04	2.22	0.18	1	0.35	1.17	0.44	1.05	0.47	2.21	14	0.0439	2.2005	0.138
4	A3_4	0	0.08	-2.34	0.14	1	-0.78	-0.56	0.4	-1.78	0.43	-4.15	13	0.0011	6.9406	0.0084
5	A3_7	0	0.02	-0.89	0.13	1	-0.15	-0.56	0.4	-0.33	0.43	-0.77	13	0.4523	0.22	0.639
6	A3_8	0	0.07	-2.26	0.14	1	-0.68	-0.72	0.41	-1.54	0.43	-3.6	13	0.0033	3.5962	0.0579
7	A3_9	0	-0.01	0.73	0.14	1	0.05	0.62	0.42	0.11	0.44	0.24	13	0.8116	0	0.9991
8	A3_10	0	-0.05	0.41	0.14	1	0.44	-0.56	0.4	0.97	0.43	2.27	13	0.0409	4.2937	0.0383