Development and Validation of the Statistics Anxiety Measure

A Dissertation

Presented to the College of Education

University of Denver

In Partial Fulfillment

of the Requirements

for the Degree

Doctorate of Philosophy

---

by

Morgan S. Earp

August 2007

Advisor: Dr. Kathy E. Green

GRADUATE STUDIES
AT
THE UNIVERSITY OF DENVER

Upon the recommendation of the Dean of the College of Education this dissertation is hereby accepted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

____
Dr. Kathy E. Green

____
Vice Provost of Graduate Studies

____
Date

TABLE OF CONTENTS

Chapter		Page
I	Introduction and Literature Review	1
	Problem	14
	Purpose	15
	Research Questions	16
II	Method	19
	Phase One: Planning	22
	Phase Two: Construction	27
	Phase Three: Quantitative Evaluation	32
	Phase Four: Validation	45
III	Results	48
	Phase Two: Construction	48
	Phase Three: Quantitative Evaluation	54
	Phase Four: Validation	109
IV	Discussion	112
	Summary of Results	114
	Methodological Conclusions	123
	Limitations	124
	Conclusion	126
	Suggestions for Further Research	127
References		130
Appendix A		136
Appendix B		137
Appendix C		139
Appendix D		145
Appendix E		146
Appendix F		147
Appendix G		151

LIST OF TABLES

Table		Page
1	Scale Development Procedure	21
2	Projected Item Pool for the SAM	29
3	Pilot Sample Distribution by Race/Ethnicity	36
4	Final Administration Sample Distribution by Race/Ethnicity	39
5	Expected Correlations between the Original SAM Domains, STARS, and SATS	46
6	Expected Correlations between the EFA SAM Domains, STARS, and SATS	46
7	Total Variance Explained by Factors in Pilot Data	57
8	Pattern Matrix for Five-factor Solution (All 60 Pilot Items)	59
9	Pattern Matrix for Six Factor Solution (All 60 Pilot Items)	60
10	Final Pattern Matrix for Five-factor Solution (All 43 Pilot Items)	61
11	Pilot MIRT Item Key	62
12	Final Administration Data Missing By Variable	65
13	Final Administration Items to Keep Based on EFA Loadings for a 5 Factor Solution Using Pilot Data	67
14	Final Administration Items to Keep Based on MIRT Weighted Misfit t Scores for a 5 Factor Solution Using Pilot Data	68
15	Final Administration Items Recommended to Keep Based on Estimate of Cronbachôs Alpha if Item were Deleted for a 5 Factor Solution Using Pilot Data.	69
16	Final Administration Items Recommended to Keep Based on Hybrid Score for a 5 Factor Solution Using Pilot Data.	70
17	Confirmatory Factor Analysis Five-factor Solution Correlations and Covariances Using Final Administration Data	73

18	Final Administration Items Recommended for Keeping Based on Hybrid Score for a 4 Factor Solution Using Pilot Data	74
19	Multidimensional Item Response Theory Response Model Parameter Estimates	86
20	Multidimensional Item Response Theory Response Model Parameter Estimates Based on Item Thresholds	88
21	Internal Consistency Reliability of Four-factor Model Subscales (Cronbach's Alpha)	91
22	DIF MIRT Model: Main Effect of Gender	93
23	DIF MIRT Model: Interaction Effect of Gender by Item	94
24	DIF MIRT Model: Main Effect of Race	95
25	DIF MIRT Model: Interaction Effect of Race by Item	97
26	DIF MIRT Model: Main Effect of Age	99
27	DIF MIRT Model: Interaction Effect of Age and Items	100
28	DIF MIRT Model: Main Effect of Course Level	102
29	DIF MIRT Model: Testing the Interaction Effect of Course Level and Items	103
30	DIF MIRT Model: Main Effect of Business Course	105
31	DIF MIRT Model: Interaction Effect of Business Course and Items	106
32	Comparison of DIF MIRT Models	107
33	DIF Test Variable Correlations	108
34	Comparison of Item 7 DIF across Gender, Race, Age, Course Level, and Business Course	109
35	Correlations among SAM, SATS, and STARS Subscales and Total Scores	110

LIST OF FIGURES

Figure		Page
1	Statistics Anxiety Domains.	12
2	Pilot Sample Distribution by Department	35
3	Pilot Sample Distribution by Age	36
4	Final Administration Sample Distribution by Department.	38
5	Final Administration Sample Distribution by Age.	39
6	Pilot Data Scree Plot	56
7	Pilot Data Item Difficulty Plot.	63
8	Confirmatory Factor Analysis 5 Factor Solution Path Diagram of R Estimates Based on Hybrid Keep Indicator Using Final Administration Data.	72
9	Confirmatory Factor Analysis 4 Factor Solution Path Diagram of R Estimates Based on Hybrid Keep Indicator Using Final Administration Data	75
10	Confirmatory Factor Analysis 5 Factor Solution Path Diagram of *t* Values Based on Hybrid Keep Indicator using Final Administration Data.	77
11	Confirmatory Factor Analysis 4 Factor Solution Path Diagram of *t* Values Based on Hybrid Keep Indicator using Final Administration Data	78
12	Confirmatory Factor Analysis 1 Factor Solution Path Diagram of R Estimates Based on Hybrid Keep Indicator using Final Administration Data	80
13	Confirmatory Factor Analysis Hybrid Model Path Diagram of R Estimates	82

14 Confirmatory Factor Analysis for the Hybrid Structural Model: Path Diagram of R Estimates 83

15 Final Administration Data Four-factor Item Difficulty Plot 90

Acknowledgements

I would like to thank my research assistant, Earl B. Earp III, for always showing interest in my topic and making time for all my dissertation needs. I would like to thank Terry Dalton for allowing me to administer my measure in all of her statistics classes at the University of Denver (DU). I would like to thank the statistics students and professors at DU for taking the time to participate in my study. I would like to thank Jann and Phelan Earp, Dulcenella Cabus, Dr. Nick Cutforth, Dr. Barb Gerhardt, Annette Nunez, and Dr. Don Bacon for their support and encouragement during the dissertation process.

I would especially like to thank my advisor, Dr. Kathy E. Green, whose encouragement, enthusiasm, guidance, and insight were invaluable in the creation and completion of this dissertation.

CHAPTER 1

INTRODUCTION AND LITERATURE REVEIW

ñThere are three kinds of lies: lies, damned lies, and statisticsò (Twain, 1924, p. 246). Over the years statistics have earned a negative reputation and for good reason. Results of numerous studies are reported on the news, on the radio, and in the paper. These results are often misinterpreted by the media and sometimes reported as fact, but are they? Correlations are interpreted as causes, and the truth may be obscured. Deming (1900-1993), an American statistician, is quoted as saying ñIn God we trust, all others bring dataò (Value Quotes, p.1) However, the problem of finding trustworthy results lies in understanding how data are collected, analyzed, and reported. If the data are confounded so are the results. Statistical illiteracy prevents the public from questioning theory, and thus theory can be misinterpreted as fact (Jordan & Haines, 2004; Schield, 2004). The general population likely lacks the skills to think critically about the results of statistically complex empirical studies. Therefore, it is essential that efforts be made to improve statistical literacy. One means for doing this is emphasizing statistical literacy in post secondary education.

Over the years researchers have found that less attention is given to statistical literacy than to literacy in general (Jordan & Haines, 2003). For example, most college students enrolled at a liberal arts school are required to take freshmen English. However, not every student is required to take statistics during their freshman year. Statistical literacy is central to making informed decisions (Jordan & Haines, 2003; Schield, 2004). Our government, economy, education, and health care system are dependent in part upon the statistical results of scientific studies. Without a clear understanding of the procedures for collecting, reporting, and analyzing data, people are misinformed. Schield stated ñé statistics are contextual; they depend on what is taken into accountò (p. 17). A clear understanding of how statistics are computed directly impacts a personôs ability to correctly interpret scientific studies (Schield). Since most of our society is dependent on statistical studies one could argue that there should be more importance placed on statistics in post secondary education.

A problem with requiring all students to take statistics in post secondary school is the feeling of ñanxietyò evoked. While statements of anxiety as it pertains to writing are met with encouragement, statements of anxiety that pertain to statistics are met with empathy (Jordan & Haines, 2003). Jordan and Haines stated ñWhile statements like óI am not good at writingô are typically answered with encouragement and reassurance that good writing can be developed through training and practice, statements like óIôm just not good at mathô are all too often answered with silence and or a sympathetic nodò (p. 17). Such empathy has worked against statistical literacy, giving rise to the notion that although everyone can learn to write, not everyone can learn statistics, thus resulting in a society in which many people are anxious of statistics.

Jordan and Haines (2003) describe quantitative literacy as having three components: foundation and mathematical skills, quantitative reasoning skills, and positive confident attitudes and beliefs about mathematics and quantitative reasoning. The last component can be seen as a result of the first two. Students lacking a foundation in math and quantitative reasoning may be more likely to develop negative attitudes and beliefs about mathematics and the quantitative reasoning involved in statistics. This may result in a lack of self confidence in situations involving math and quantitative reasoning. Identifying individuals lacking foundational skills and holding negative attitudes is essential to creating general quantitative literacy.

In the past twenty years, several studies using scales measuring attitudes toward statistics and statistics anxiety found strong correlations between attitudes towards statistics, statistics anxiety, and statistics performance. At least four measures exist for assessing attitudes towards statistics. These measures include the Statistics Attitude Scale (Cruise, Cash, & Bolton, 1985), Statistics Attitude Survey (Roberts & Bilderback, 1980), Survey of Attitudes Toward Statistics (Schau, Stevens, Dauphinee, & Vecchio, 1995), and the Attitude Toward Statistics scale (Wise, 1985). Roberts, Bilderback, and Saxe (1982) argued that attitudes toward statistics play a key role in student success in statistics classes. In 1980, Roberts and Bilderback published the Statistics Attitude Survey (SAS), with the main purpose of testing whether it was an effective measure in the prediction of performance in statistics. Five years later, Wise published the Attitude Toward Statistics (ATS) scale, arguing that the SAS was an invalid measure of statistics attitudes since ñé at least a third of the measure dealt with such topics as student success in solving statistics problems or success in understanding statistics conceptsò (pp. 401-402). Wise claimed that items on the SAS ñé appeared to be measuring student achievement rather than student attitudeé ò and that items were ñé inappropriate for students who are just beginning their statistics courseò (p. 402). As a result, Wise developed the ATS to measure attitudes toward statistics in introductory statistics courses. He claimed that the construct of attitudes toward statistics is comprised of two factors: attitude toward the course and attitude toward the field. Ten years later, Schau et al. (1995) developed the Survey of Attitudes Toward Statistics (SATS), arguing that attitude toward statistics was made up of the following four factors: affect, cognitive competence, value, and difficulty. The Statistics Attitude Survey (SAS), Attitude Toward Statistics (ATS), and the Survey of Attitudes Toward Statistics (SATS) have all been shown to be reliable measures of attitudes toward statistics; however, researchers disagree on the factors comprising statistics anxiety or attitude.

Attitudes toward statistics and statistics anxiety have been found to be highly correlated, with attitude toward statistics often influencing statistics anxiety (Mji & Onwuegbuzie, 2004; Zeidner, 1991). Students with negative experiences from mathematical or statistical courses or instructors are often scared and carry such memories in the form of anxiety. Students with negative attitudes toward statistics are thought to be highly anxious with regard to statistics (Mji & Onwuegbuzie). At least three reliable and valid measures of statistics anxiety exist: the Statistics Anxiety Rating Scale (Cruise & Wilkins, 1980), the Statistics Anxiety Scale (Pretorius & Norman, 1992), and the Statistics Anxiety Inventory (Zeidner, 1991). Zeinder defined statistics anxiety as, ñé a particular form of performance anxiety characterized by extensive worry, intrusive thoughts, mental disorganizations, tension, and physiological arousalò (p. 319). Mji and Onwuegbuzie reported that 66% to 80% of all graduate students report unmanageable levels of statistics anxiety. They concluded that this form of anxiety is a situation-specific trait as students are expected only to experience it when studying or applying statistics.

Cruise, Cash, and Bolton (1985) published the Statistics Anxiety Rating Scale (STARS), arguing that statistics anxiety is comprised of the following six factors: perceived worth of statistics, interpretation anxiety, test and class anxiety, conceptual self-concept, fear of asking for help, and fear of statistics teachers. Six years later Zeidner (1991) developed the Statistics Anxiety Inventory (SAI), arguing that statistics anxiety is comprised of the following two factors: statistics content anxiety and statistics test anxiety. Zeidner pointed out that statistics anxiety is conceptually different from test anxiety, since statistics anxiety ñé includes a personôs reaction to statistics content as well as performance evaluationö (p. 319). A personôs reaction to statistics may be partly due to their attitude toward statistics in general.

Statistics anxiety and attitude toward statistics are correlated. The two constructs share many characteristics but neither construct is clearly delineated. Published measures of statistics anxiety and attitude toward statistics have been found to be both reliable and valid, but the research thus far focuses on differentiating them instead of focusing on what they have in common. For example, Schau et al. (1995) developed the SATS and used the ATS for validation. They found that the ATS course scale correlated positively and significantly with all four SATS scales: Affect = .79, Cognitive Competence = .76, Value = .40 and Difficulty = .42. In addition the ATS field scale correlated positively and significantly with three of the SATS scales: Affect = .34, Cognitive Competence = .38, and Value = .76. Roberts and Reese (1987) defended the SAS against criticism from Wise (1985) by correlating the SAS with Wiseôs measure (ATS), and found the correlation between the SAS and ATS was .88. Mji and Onwuegbuzie (1987) tested construct validity by correlating the STARS with the ATS, hypothesizing a link between statistics anxiety and attitudes toward statistics. They found the following:

> é after applying the Bonferroni adjustment (Ŭ=.05/6=.0083), the Attitudes Toward the Course subscale scores were statistically significant and negatively correlated with five of the six anxiety dimensions, as follows: Worth of Statistics ($r = -.65$, $p < .0083$), Interpretation Anxiety ($r = -.25$, $p < .0083$), Test and Class Anxiety ($r = -.63$, $p < .0083$), Computational Self-Concept ($r = -.59$, $p < .0083$), Fear of Asking for Help ($r = -.001$, $p < .0083$), and Fear of Statistics Teachers ($r = -.30$, $p < .0083$)(p. 242).

Based on the cited results, these measures are clearly tapping into strongly correlated constructs, if not the same construct. Arguments among scholars center on which set of highly correlated measures is the more reliable and valid continue. It is important that we take a step back and examine what the research has in common thus far.

According to Zeidner (1991), statistics anxiety is related to statistics test anxiety, statistics content anxiety, and an individualÕs history of success and failure experiences in situations involving math. Math anxiety is influenced by math test anxiety, numerical anxiety (Rounds & Hendel, 1980), and lack of an adequate foundation in math (Burton & Russell, 1979). Statistics anxiety is a special case of performance anxiety which is related to extensive worry, intrusive thoughts, mental disorganization, tension, psychological arousal, and instructional situations (Zeidner). Statistics anxiety influences an individualÕs level of performance in college statistics and his/her tendency to avoid classes involving statistics (Zeidner).

According to Benson (1987) and Benson and Bandalos (1989) statistics anxiety is related to test anxiety, math anxiety, and an individualÕs history of success and failure experiences in situations involving math. An individualÕs history of success and failure experiences in situations involving math are influenced by affective, cognitive, and social factors (Sarason, 1980; Zeidner & Safir, 1989).   Attitude, cognitive, and social factors are influenced by unrealistically high paternal peer pressures and expectations for success in math studies, high expectations of punishment for failure to meet demands in mathematic solving situations, unfortunate past personal experiences with math course material (teaching environment, instructors etc.), an individualǿs low level of math reasoning ability and concomitant feelings of helplessness and despair in solving mathematical problems, a low degree of academic self confidence in general, and a perceived low probability of success in statistics (Sarason; Zeidner & Safir). Fishbein and Ajzen (1975) found that attitude influences behavior intentions, which are also influenced by subjective norms, and later transcend into active behavior. In addition, Eagly and Chakin (1992) concluded that attitude influences steps in information processing, future behavior, achievement, and motivation to continue learning.

According to Hendel (1980), Tobias (1987), and Richardson and Woolfolk (1980), statistics anxiety is related to personal background (ethnicity, social class, and gender), prior educational experience (success and failure in elementary and high school math and number and quality of prior math classes), and learning motivation, which is influenced by quality of instruction in statistics, personality and attitude of instructor, and difficulty of the material and test procedures. They also concluded that personal background, prior educational experience, and learning motivation variables influence course performance and student learning. Cruise and Wilken (1980) concluded that statistics anxiety is influenced by test and class anxiety, interpretation anxiety, fear of asking for help, fear of statistics teachers, perceived worth of statistics, and conceptual self concept. Aiken (1976) found that math anxiety and math attitudes affect overall performance in math. In addition, Sutarso (1992) concluded that statistic performance is influenced by anxiety, computer and math skills, and statistical pre knowledge. Lastly, Smith (1981) concluded low math self-esteem leads to math anxiety.

Over the last few decades the sophistication of statistical analysis has greatly increased with the widespread use of computers. With such advances, performing complex statistical operations is as easy as pointing and clicking. However, despite the accessibility of complex statistical software programs, statistical literacy has not been addressed until recent years. The repercussions of the widening gap between software development and statistical literacy are catching up with us faster than we can address the problem of statistical illiteracy. The availability of software has resulted in an increase in the number of statistical studies as well as students required to interpret them. The problem is the number of people becoming statistically literate has most likely not been proportional to the number of people performing and interpreting statistics. Today it is more important than ever that research is conducted to remedy the problem, because without a clear understanding of what promotes today's phenomenon of statistical illiteracy, the problem is sure to worsen.

In a society where statistics are generally respected but relatively rarely understood, statistical literacy can make a difference in one's quality of life. However, there is more to becoming statistically literate than just taking a required statistics course, hoping to pass, and then forgetting it. For those people who lack a foundation in mathematical and quantitative reasoning, gaining the self confidence to believe they can learn statistics and the desire to do so can pose a problem (Jordan & Haines, 2003). One way to address this problem is to identify those individuals who suffer from statistics anxiety and gain a better understanding of what specifically contributes to their anxiety. A number of measures discussed above were created specifically for this reason. However, instead of working together and building models to address and understand anxiety towards statistics, the measures are based on competitive theories. The competing theories are interrelated and thus are often used to argue for mutual measure validity. This review and subsequent new measure bears a focus on what unites the theories and extant measures, with the goal of creating a measure of statistics anxiety and attitudes towards statistics that integrates prior work and addresses the problem of statistical illiteracy.

Competing factors contributing to statistical anxiety and attitudes toward statistics were diagrammed and placed in a single diagram, which was then visually analyzed for similarities across constructs in Figure 1. Thematic analysis of underlying theoretical traits of statistics anxiety revealed the variables represented in Figure 1. These variables were categorized into six common domains using an inductive analysis to discover patterns, themes, and ultimately create categories based on the previous research surrounding statistics anxiety and attitudes toward statistics. The six categories identified though inductive analyses which are later referred to as domains include: anxiety, fearful behavior, attitude, expectations, history and self-concept, and performance.

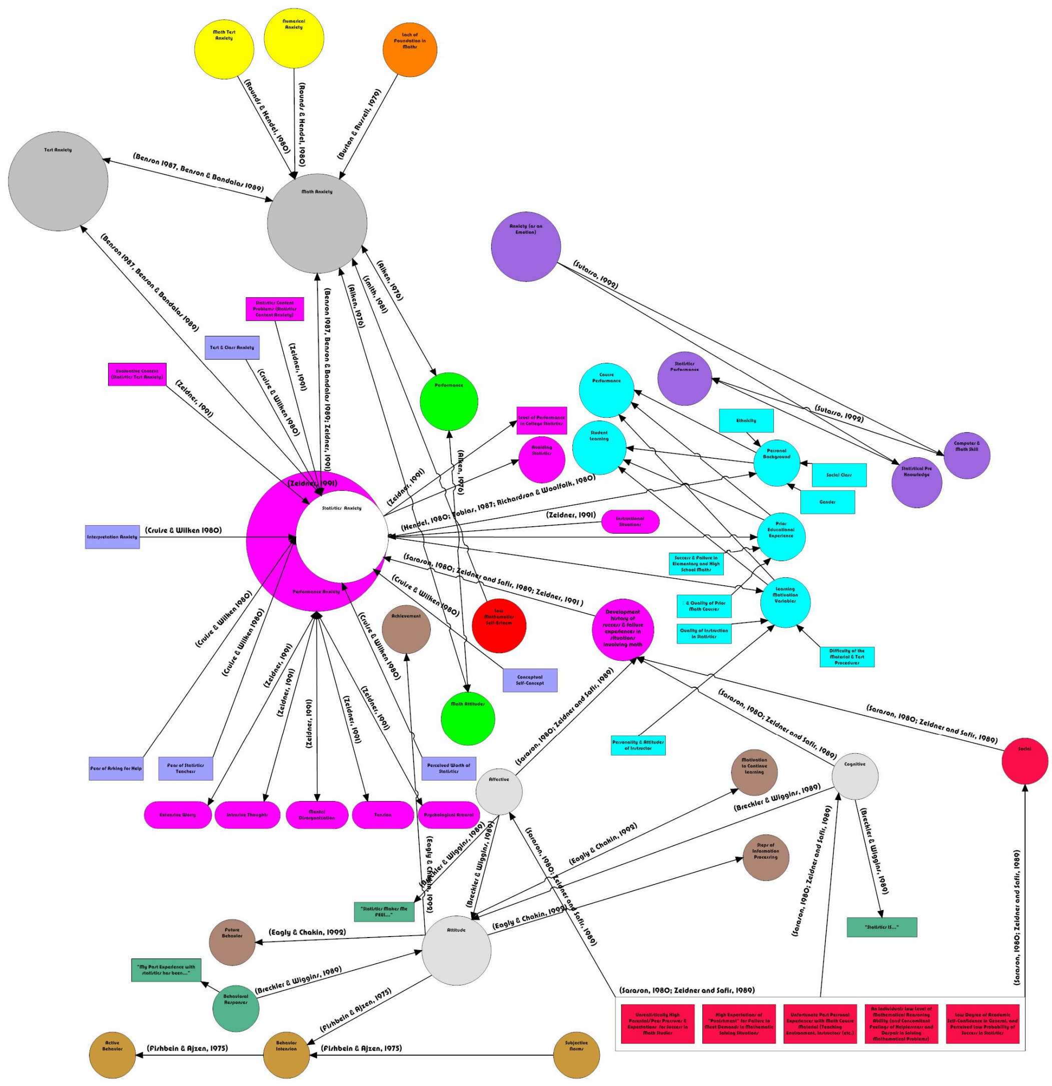

Figure 1. Statistics Anxiety Domains.

Six domains were found that encompassed all the previous constructs of statistical anxiety and attitudes toward statistics. These domains were identified through analyses of different theories related to statistics anxiety and attitudes toward statistics (Aiken, 1976; Benson, 1987; Benson & Bandalos, 1989; Breckler & Wiggins, 1989; Burton & Russell, 1979; Cruise & Wilken, 1980; Eagly & Chakin, 1992; Fishbein & Ajzen, 1975; Hendel, 1980; Richardson & Woolfolk, 1980; Rounds & Hendel, 1980; Smith, 1981; Sarason, 1980; Sutarso, 1992; Tobias, 1987; Zeidner, 1991; Zeidner & Safir, 1989).

The anxiety domain is comprised of statistics content anxiety, statistics test anxiety (Zeidner, 1991), class anxiety, interpretation anxiety (Cruise & Wilken, 1980), test anxiety, math anxiety (Benson 1987; Benson & Bandalos 1989), math test anxiety, numerical anxiety (Rounds & Hendel, 1980), and lack of math foundation (Burton & Russell, 1979).

The fearful behavior domain is comprised of items that address fear of asking for help, fear of statistics teachers (Cruise & Wilken, 1980), extensive worry, intrusive thoughts, mental disorganization, tension (Zeidner, 1991), and behavioral responses (Breckler & Wiggins, 1989).

The attitude domain is comprised of items that address math attitudes (Aiken, 1976), perceived worth of statistics (Cruise & Wilken, 1980), affect (Sarason, 1980; Zeidner & Safir, 1989), and psychological arousal (Zeidner, 1991).

The expectation domain is comprised of items that address subjective norms (Fishbein & Ajzen, 1975), motivation to continue learning, steps in information processing (Eagly & Chakin, 1992), cognition (Breckler & Wiggins, 1989), social expectations, unrealistically high parental/peer pressures to succeed in math, high expectations of punishment for failure to meet demands in mathematical solving situations, unfortunate past experience with math course material, and low level of math reasoning ability (Sarason, 1980; Zeidner & Safir, 1989).

The history and self-concept domain is comprised of low mathematics self esteem (Smith, 1981), history of success and failure experiences in situations involving math (Sarason, 1980; Zeidner & Safir, 1989), self-concept (Cruise & Wilken, 1980), prior educational experience, motivation to learn (Hendel, 1980; Richardson & Woolfolk, 1980; Tobias, 1987), and instructional situations (Zeidner, 1991).

The performance domain is comprised of self-perception of ability to perform in statistics.

The identified six domains are, to some extent, overlapping, but the extent of their overlap is an empirical issue.

Problem

Statistical literacy can make a difference in a personǒs quality of life including everything from decisions regarding health, politics, and economics to employability; therefore, it is essential that students not only gain a strong foundation in statistics, but also that they develop a strong self-concept when it comes to interpretation of numerical data (Jordan & Haines, 2003). Identifying individuals suffering from statistics anxiety and gaining a better understanding of the domains that contribute to such anxiety is a start to addressing the problem of statistical illiteracy today. A number of studies pertaining to statistics anxiety and attitudes toward statistics have been conducted, but instead of working together to understand and address these problems, the majority of studies are working against each other. Each claims that their theory is an advance over those of previous authors. Therefore, the researcher aimed to create a measure of statistics anxiety that unites the literature thus far in the aims of truly understanding the problem and that will allow researchers to address the problem of statistical illiteracy in the future.

Purpose

Rather than disputing the previous research on statistics anxiety and attitudes toward statistics, the purpose of this study was to determine what the research has in common and construct an instrument that assesses multiple dimensions of statistics anxiety. Using the diagram in Figure 1, variables cited in previous research on statistics anxiety and attitudes toward statistics were visually analyzed, and the following six domains were identified by combining a number of theories: anxiety, fearful behavior, attitude, expectation, history and self-concept, and performance. The Statistics Anxiety Measure (SAM) is expected to measure statistics anxiety by tapping into each of the six domains. As the purpose of this study was to develop and validate a new multidimensional measure, attention was given to the measurement theory followed in measure development. There are two competing methods for assessing validity of measures: classical test theory and item response theory (IRT). Instead of choosing between the two, the researcher used both methods to assess domain and item fit.

IRT offers several benefits over classical test theory: adapting measures and instruction without norm referencing, assessing reliability of data obtained from persons and items, predicting responses based on person traits or persons based on responses, deleting person scores with no variability or consistency, and assessing whether items reflect person ability/attitude. Classical test theory assumes Likert-type scales are interval when they are not, item difficulty is sample dependent, person ability depends on difficulty of test items, and provides a single standard error of measurement representing all raw values. IRT generates an interval scale, item difficulty is independent of sample ability, person ability is independent of test difficulty, and standard error estimates exist for all raw scores. Most current applications of IRT assume measure unidimensionality, though interest is increasing in application of multidimensional IRT models. In the present study, I was interested in measuring statistics anxiety with major assumptions (e.g., assumptions regarding dimensionality) upheld, and thus I explored this newly developed technique of multidimensional IRT further. I was also interested in comparing multidimensional IRT and classical test theory multidimensional assessment approaches.

In order to test the structure of the SAM, the researcher used both classical test theory and IRT. Traditionally, assessment of multidimensional models was limited to classical test theory, but today new avenues have arisen. Currently psychometricians assess measures in two fashions, using either classical test theory or IRT. Classical test theory easily assesses single or multiple factors, whereas IRT traditionally only assesses single factor models. In recent years, psychometricians have developed methods for assessing multidimensional constructs using IRT.

<u>Research Questions:</u>

Do statistics anxiety items generated to reflect the six identified domains (anxiety, fearful behavior, attitude, expectations, history and self-concept, and performance) factor appropriately into the six domains?

Does the Statistics Anxiety Measure evidence adequate reliability and validity?

Does multidimensional item response theory provide better assessment of this multidimensional measure than classical test theory?

Does the SAM significantly correlate with the STARS and the SATS, more specifically does the STARS correlate more highly with the anxiety domain of SAM, and does the SATS correlate more highly with the attitude domain of SAM?

<u>Definitions:</u>

Quantitative Literacy: Jordan and Heines (2003) define quantitative literacy as contextually appropriate decision making, understanding data and statistical inference, ability to interpret data, to think critically about public issues, to use computers, and to understand and generate graphs and statistics.

Statistical Literacy:

> Statistical literacy includes many elements of quantitative literacy. It involves the mathematical approach in focusing more on signal and pattern than on noise or chance; it involves the statistical approach in focusing on the role of context, conditional reasoning, and variation. But statistical literacy goes beyond quantitative literacy or numeracy by focusing on the ability to read, to interpret, and to communicate. Numeracy focuses primarily on numbers; statistical literacy focuses more on the words framing the numbers. (Schield, 2004, p. 16)

<u>Delimitations:</u>

Through broad sampling at the University of Denver (DU) across discipline and level, the researcher expected final results to be generalizable to the DU population; however without broadening the study to other populations outside of DU there is no assurance that the results will be generalizable to the population of statistics students as a whole.

CHAPTER 2

METHOD

Introduction

In this chapter, the procedures that were followed in the development and validation of the Statistics Anxiety Measure (SAM) are discussed. This chapter starts with an overview of the design followed by a description of the four phases used to develop a valid measure of statistics anxiety: Phase 1: Planning, Phase 2: Construction, Phase 3: Quantitative Evaluation, and Phase 4: Validation.

Study Design and Purposes

According to Jordan and Haines (2003) quantitative literacy is partly determined by attitudes towards math and quantitative reasoning as well as self confidence in oneǒs math and quantitative reasoning skills. Statistical literacy includes these elements but goes beyond quantitative literacy to also include the ability to interpret and communicate the words framing numbers (Schield, 2004). One of the major obstacles to obtaining statistical literacy lies in statistics anxiety. When someone is uneasy in situations involving statistics, it can hinder their ability to learn. Therefore, as stated previously, one part of addressing statistical illiteracy is addressing statistics anxiety. By identifying people who suffer from statistics anxiety and understanding where that anxiety stems from, we have the power to address it.

There are a number of measures for statistics anxiety and attitudes toward statistics; however the focus thus far has been to differentiate them as opposed to uniting them. In the current study, the Statistics Anxiety Measure (SAM) aimed to make a comprehensive construct that incorporated all of the theories found, thus relating statistics anxiety and attitudes toward statistics. The SAM assists in identifying students who initially suffer from statistics anxiety in statistics-related courses. There were two purposes of this study: (1) to unite the literature thus far by creating items in six domains that fully covered the previous theories regarding statistics anxiety, and (2) to compare the assessmentôs psychometric properties using both classical test theory and multidimensional item response theory.

The SAM assessed the following domains: anxiety, performance, history and self-concept, expectations, attitude, and fearful behavior. The SAM was intended to measure six subscales of statistics anxiety which made up the full pilot scale. The SAM also included a demographic section, and was analyzed at the individual level. There were two purposes of including demographics in the measure: (1) to assess whether statistics anxiety was more prevalent among certain groups, and (2) to assess how representative the sample was of the actual University of Denver (DU) population.

The SAM was developed in four phases based on the recommendations provided by Benson and Clark (1982) and DeVellis (2003) as well as the steps provided in Cox et al. (2006). Table 1 provides the scale development procedure for the SAM.

Table 1

Scale Development Procedure

Development Phase	**Scale Development Steps**
Planning	- Identified the purpose of SAM - Identified the audience that the results of the SAM study will be most important to: statistics students, professors, and persons in charge of curriculum development - Conducted a literature review in which the competing theories of statistics anxiety and attitudes toward statistics are united - Conducted two pilot studies to try out different potential items - Conducted two interviews with students taking introductory statistics - Selected a four point rating scale as the format for item response
Construction	- Determined and defined domains by linking visual diagrams of various theories related to statistics anxiety and attitudes towards statistics - Generated an item pool with items that are distinguishable both by domain and level of agreeability - Conducted expert reviews of all items for content and

	agreeability validation - Reduced item pool based on feedback from the expert reviews -Conducted two cognitive interviews and incorporated feedback
Quantitative Evaluation	- Piloted new items in order to conduct item analysis and analysis of structure via exploratory factor analysis, confirmatory factor analysis, and multidimensional item response theory - Reduced item pool to only the most valid and reliable items and factors in terms of domain and overall measure fit - Conducted a field administration of reduced items and factors to a second sample and assessed reliability of the refined SAM. - Verified subscales using confirmatory factor analysis as well as multidimensional item response theory - Assessed the internal consistency reliability of the six subscales using both classical test theory and multidimensional item response theory - Assessed patterns of response
Validation	- Assessed convergent validity. - Assessed differences based on demographic groupings

Phase 1: Planning

Phase 1 focused on measure development including the purpose, range, and population the SAM was intended for. The following steps were taken to develop the SAM: a literature review, two pilot studies, two interviews, inductive analysis of the literature, and review of item structure and placement. The literature review is provided in chapter one along with the inductive analysis of the literature which was displayed in Figure 1.

Pilot Studies

The two pilot studies were carried out in 2003. The first pilot study was conducted in an undergraduate level ñIntroduction to Statisticsò course at the University of Denver in the Danielôs College of Business (DCB) using the Fear of Statistics Test (FST). The FST was the precursor to the SAM and was originally developed in an introductory psychometric theory class by the researcher and two fellow doctoral students. The researcher conducted an initial pilot study in a class consisting of thirty-four students (21 males and 13 females). The majority of students were white/Caucasian (88.2%), males (61.8%), ages twenty to twenty-one (64.8%), majoring in business sectors (49%), with a GPA greater than 3.0 (79%). The FST consisted of 6 demographic, 10 four-point rating scale items measuring fear of language, 10 four-point rating scale items measuring fear of numbers, and 10 four-point rating scale items measuring fear of outcome in the course, for a total of 36 items.

Due to the limited sample size (n=34), data were recollected during Spring Quarter 2003 from the following classes: a ñStatistical Methods Iò class in the Graduate School of International Studies (GSIS), and four undergraduate level ñIntroduction to Statisticsò courses in DCB. The majority of students sampled were white/Caucasian (77.2%), ages nineteen to twenty-one (67.6%), majoring in business sectors (57.7%), with an average of GPA of 3.42. The FST consisted of 6 demographic items in the first section, 6 four-point rating scale items measuring fear of language, 8 four-point rating scale items measuring fear of numbers, and 6 four-point rating scale items measuring fear of outcome in the class.

In the first pilot study the researcher used a sample of 31 cases to conduct an exploratory factor analysis (EFA) with the original 30 items of the FST as indicators. To test the number of factors indicated by the items, the researcher examined a scree plot of the eigenvalues resulting from the analysis. The scree plot demonstrated that a one-factor solution could fit the data. The plot also indicated that a four-factor solution might also fit the data. Therefore, the researcher chose to examine both the one-factor and the four-factor solutions. Since factors were assumed to be correlated, oblique rotation was used in order to determine factor loadings. Due to cross loadings, items loading under .50 were eliminated before conducting the second pilot study.

In the second pilot study, the researcher used results from the first pilot study to theorize the following constructs for the FST: Factor 1; anxiety as it relates to using calculators/computers and tables, Factor 2; the degree to which students feel they struggle with statistics, Factor 3; fear as it relates to memorizing formulas necessary to do well in introductory statistics, and Factor 4; the degree to which student attributes logic skills to understanding statistics. Since Factor 4 only had one indicator it was not tested in the second pilot study.

For the second pilot study, the researcher used a sample consisting of 152 cases. Ten items were eliminated based on the results of the first pilot study, since they had intercorrelations greater than .50 and were considered to have poor fit within their assigned domain. A confirmatory factor analysis (CFA) was used to analyze the remaining 20 four-point rating scale items. Because the data were ordinal, the researcher used PRELIS to calculate the asymptotic covariances and polychloric correlations of all 20 items. Using LISREL with weighted least square estimation for the models implied by the EFA, a three-factor solution was found in the CFA.

Due to the facets having similar meanings, the researcher first tested a one-factor solution, and found that the misfit was greater for a one-factor model than for either a three- or four-factor CFA model. Therefore, the researcher concluded that the three-factor model best fit the data. The FST factor structure was ultimately driven by the solution; however, according to Klein (1998) the best factor structures are driven by theory, not empirical solutions. Therefore, the researcher turned to the current literature surrounding fear of statistics to build a more reliable and valid measure.

Student Interviews

The first two pilot studies were conducted prior to a thorough literature review. Although much of what was contained in the FST was supported by the literature, the researcher found the term fear was incorrectly used. Fear pertains to situations that involve the perception of life or death, where anxiety pertains to all other stressful situations (Nemours Foundation, 2002). Therefore, interviews were conducted with two students in a graduate level ñIntroduction to Statisticsò course during Spring Quarter of 2005. The interviews began with one main question ñWhat is statistics anxiety and how does it differ from fear of statistics?ò The main question led to two hour-long discussions with each student about what contributed to statistics anxiety and fear of statistics. It was clear from the interviews that the two concepts were distinguished by the life and death factor. However, since death is not usually an outcome of failing statistics, the researcher used the findings obtained in interviews to begin theoretical investigations into statistics anxiety to supplement the information gained from the pilot studies.

Response Format

The last piece of phase one laid out the SAM response format. To ensure that respondents provided decisive responses the SAM used a four-point rating scale as opposed to a Likert scale. Once the SAM item pool was fully constructed, it consisted of approximately three times as many items as necessary for measuring each domain (120 items = 20 items per factor) (DeVellis, 2003). Upon expert review, the number of items was reduced by half in order to ensure the burden on participants was low and to ensure that only the most useful items were included in the SAM.

Phase 2: Construction

Phase two was the construction of SAM. This section laid out the process for both generating items that measure each of the six domains as well as the process of item elimination based on expert review. There were four subsections of Phase Two: item pool creation, expert review, item selection, and cognitive interviews.

Item Pool Creation

The content developed for the SAM was derived from a number of previously discussed statistics anxiety and attitude toward statistics theories (Aiken, 1976; Benson, 1987; Benson & Bandalas, 1989; Breckler & Wiggins, 1989; Burton & Russell, 1979; Cruise & Wilken, 1980; Eagly & Chakin, 1992; Fishbein & Ajzen, 1975; Hendel, 1980; Richardson & Woolfolk, 1980; Rounds & Hendel, 1980; Smith, 1981; Sarason, 1980; Sutarso, 1992; Tobias, 1987; Zeidner, 1991; Zeidner & Safir, 1989). A number of theories were deconstructed; visual analysis suggested six domains of statistics anxiety: anxiety, fearful behavior, attitude, expectations, history and self-concept, and performance. Items were developed to measure the six domains. Items were also adapted from the FST.

According to Benson and Clark (1982), the first step in item pool creation is defining the overall construct and the domains which comprise it. The objective of the SAM and the definitions of the six domains are as follows:

Objective: The SAM was intended to identify students who suffer from statistics anxiety and more specifically which domain of statistics anxiety is most concerning for them.

Domain Definitions:

Anxiety: The anxiety domain was defined as anxiety as it relates to tests, math, the class, statistics content, numbers, and interpreting numerical data. Persons scoring high in this domain will feel anxious in class situations involving tests, math, numbers, statistics, and/or in the process of interpreting results derived from empirical analysis.

Performance: The performance domain was defined by self-reported perception of course performance, ability to perform statistical operations, and ability to learn statistical concepts.

History and Self-Concept: The history and self-concept domain was defined by developmental history of success and failure in situations involving math, low math self esteem, low self-concept, prior educational experiences with math in elementary and high school, perceived quality of prior math classes, motivation to learn, difficulty of material in previous math classes, and quality of instruction in previous math and statistics classes.

Expectations: The expectations domain was defined by social and cognitive expectations, unrealistically high expectations from parents and/or peers, and high expectation of punishment.

Attitude: The attitude domain was defined by attitude as it pertains to math, worth of statistics, and psychological arousal with respect to the level of personal fulfillment gained in the practice of statistics.

Fearful Behavior: The fearful behavior domain was defined by extensive worry, intrusive thoughts, mental disorganization, tension, and fear as it relates to instructors, asking for help, past behavior, current behavior, and future behavior.

Once the construct and domains were defined, the researcher planned the magnitude of the item pool as it pertained to expert review, piloting, field administration, and validation. The table of specifications is provided as Table 2.

Table 2

Projected Item Pool for the SAM

Domain	Item Pool	Pilot	Field Administration
Demographics	6	6	6
Anxiety	20	10	5
Performance	20	10	5
History and Self-Concept	20	10	5
Expectations	20	10	5
Attitude	20	10	5
Fearful Behavior	20	10	5

Items developed for the SAM were generated from findings in the literature review, items that functioned well on the FST, and the thematic analysis carried out and presented as Figure 1.

Expert Review

According to DeVellis (2003) one should enlist between 6 and 10 experts on the measure content to review items for a newly constructed test. The expert panel was asked to complete a survey rating of the quality of each item, appropriateness for the domain, and the perceived level of agreeability for that item using an ordinal scale. The goal was that each item was of high quality and valid, and that the items within each domain followed an ordinal scale in terms of agreeability. Furthermore, experts were asked to rate domain items on agreeability so that items in each domain were likely to yield a continuum. Items measuring anxiety, fearful behavior, attitude, expectations, history and self concept, and performance were reviewed by five experts. The five experts enlisted for this study included three professors at the University of Denver, one person in charge of assessment for her division at the University of Denver, and the chief cognitive research Methodologist for the United States Department of Agricultureøs (USDA) National Agricultural Statistics Service (NASS). The three professors asked for their expertise were the program chair for a quantitative research program for a college of education, an assistant professor who teaches survey design, and a professor teaching statistics in a school of social work. The person in charge of assessment for her division asked for her expertise was a coordinator of experience for student life. This person was in charge of assessment and has created numerous scales to measure effectiveness of student life programs. The Chief Cognitive Research Methodologist for NASS is responsible for in-house survey methodology training as well as item content review of NASS surveys including cognitive interviews etc.

Experts were emailed asking for their participation in the review of SAM. The email laid out the purpose of the study and the requirements of participation: complete a Likert-scale rating of the quality of each item, appropriateness for the domain, and the perceived level of agreeability for that item using an ordinal scale. Experts were contacted at the end of December, 2006 and were provided with an online link to perform the above tasks. Using an online process was expected to reduce the burden and confusion of participating in the expert review, as well make results of the review available immediately to the researcher upon completion. Two weeks after initial contact and request, the researcher sent out a reminder to experts who had yet to respond and a thank you letter to those who had. It was expected that responses would be received from at least five experts.

Item Selection

Once results of the expert review were obtained, the researcher began the process of item clarification and elimination. Items were rated for quality using a Likert scale. Items with mean ratings lower than two were assessed for possible improvement or dropped from the SAM. Items were rated for appropriateness of domain, and results were analyzed to determine if certain items were better suited for other domains included on the SAM. Items were rated using a Likert scale for agreeability. Agreeability was analyzed in each domain and considered in the selection and ordering of items for the SAM pilot.

Cognitive Interviews

The researcher conducted two cognitive interviews once the survey items were in near-final form for the pilot. Cognitive interviews were held with two students registered for introductory level statistics at DU during Winter Quarter, 2007. Both participants were asked to think aloud as they worked their way through the survey. This allowed the researcher to hear how the items were being interpreted and how responses were being decided, so that further clarification could be made where necessary.

Phase 3: Quantitative Evaluation

Purpose

Evaluation of the SAM took place in two stages: a pilot and a field administration. The pilot was used to determine the fit of items within their domains and in the overall measure. Items with low fit were eliminated and the SAM was readministered in a field administration. Items were analyzed using classical test theory for item-total correlations and internal consistency reliability within the six domains in both the pilot and the field administrations. Item fit in the pilot was assessed using SPSS (exploratory factor analysis) and ConQuest (multidimensional item response theory: Wu et al., 2003). Item fit in field administrations were assessed using both LISREL software (confirmatory factor analysis) and ConQuest software (multidimensional item response theory: Wu et al., 2003). The structure of SAM was initially tested using EFA in SPSS and ultimately tested using CFA in LISREL. Since the data assumed an ordinal scale, the researcher used PRELIS to calculate asymptotic covariances and polychloric correlations of all four-point rating scale items before assessing structure through CFA. Using weighted least square estimation in CFA, the theorized final administration model of the SAM was tested for factor structure. Item fit was assessed using multidimensional item response theory using ConQuest which makes determinations of item fit using computations originated by Wright and Masters (1982). The reliability of subscales was assessed by comparing the similarity in factor solutions between the pilot study to the field administration.

Items that demonstrated poor fit and reliability in the pilot were eliminated before the field administration. Specific domains that were not uniquely identified and items that were not distinguishable in factor structure were either eliminated or the subscales were combined to form fewer than the originally intended number of domains before the field administration took place.

Participants

Pilot. For the initial administration of SAM, fifteen professors from ten departments at the University of Denver were sent letters electronically asking for their participation in a field study of the revised SAM. The requests included an introductory letter explaining the rationale behind the study as well as the importance of it with a list of dates and times. Given agreement on the part of the professors, the pilot measure was administered during Winter Quarter to twenty classes across seven departments with a sample of 347 students (Figure 2). Five undergraduate-level statistics classes participated (n = 131) and fifteen graduate-level statistics classes participated (n = 215). Over half of the pilot sample was female (59.9%) and the majority of students sampled were Caucasian (75.8%) (Table 3). The average age of participants in the pilot study was 25.63 (SD = 7.00), and ranged from 18 to 59 (Figure 3).

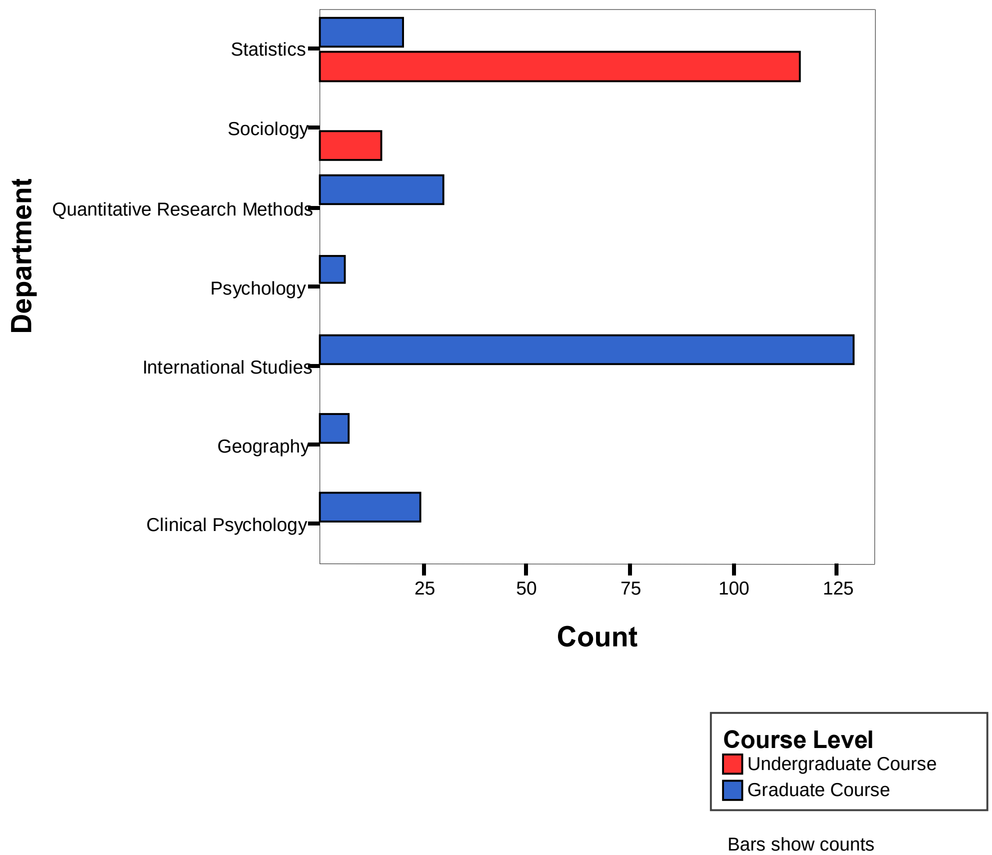

Figure 2. Pilot Sample Distribution by Department.

Table 3

Pilot Sample Distribution by Race/Ethnicity

Race

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	African American	9	2.6	2.6	2.6
	African	1	.3	.3	2.9
	Asian/Pacific Islander	35	10.1	10.1	13.0
	Middle Eastern	7	2.0	2.0	15.0
	Native American	2	.6	.6	15.6
	Caucasian/White	263	75.8	75.8	91.4
	None of the Above	20	5.8	5.8	97.1
	Jewish	1	.3	.3	97.4
	Hispanic	8	2.3	2.3	99.7
	Indian	1	.3	.3	100.0
	Total	347	100.0	100.0	

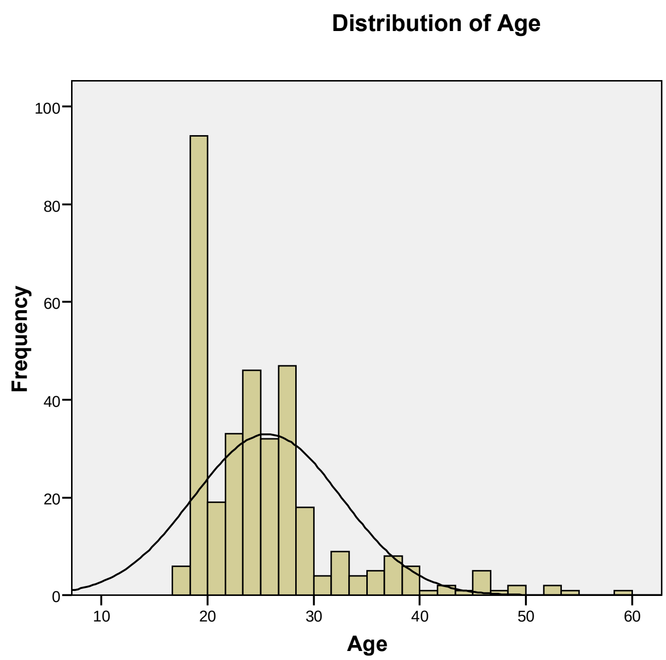

Figure 3. Pilot Sample Distribution by Age.

Final Administration. The final administration of SAM took place primarily to assess structure and scale and item functioning. For the final administration of SAM, twenty-one professors from eight departments at the University of Denver were sent letters electronically asking for their participation in a field study of the revised SAM. The requests included an introductory letter explaining the rationale behind the study as well as the importance of it and a list of dates and times. Given agreement on the part of the professors, the SAM was administered during Spring Quarter to twenty-one classes across eight departments with a sample of 433 students (Figure 4). Eight undergraduate-level statistics classes were surveyed (n = 223) and thirteen graduate-level statistics courses were surveyed (n = 207). Over half of the final administration sample was female (58.7%) and the majority of students sampled were Caucasian (80.8%) (Table 4). The average age of participants in the final administration study was 24.35 (SD = 6.51), and ranged from 18 to 55 (Figure 5).

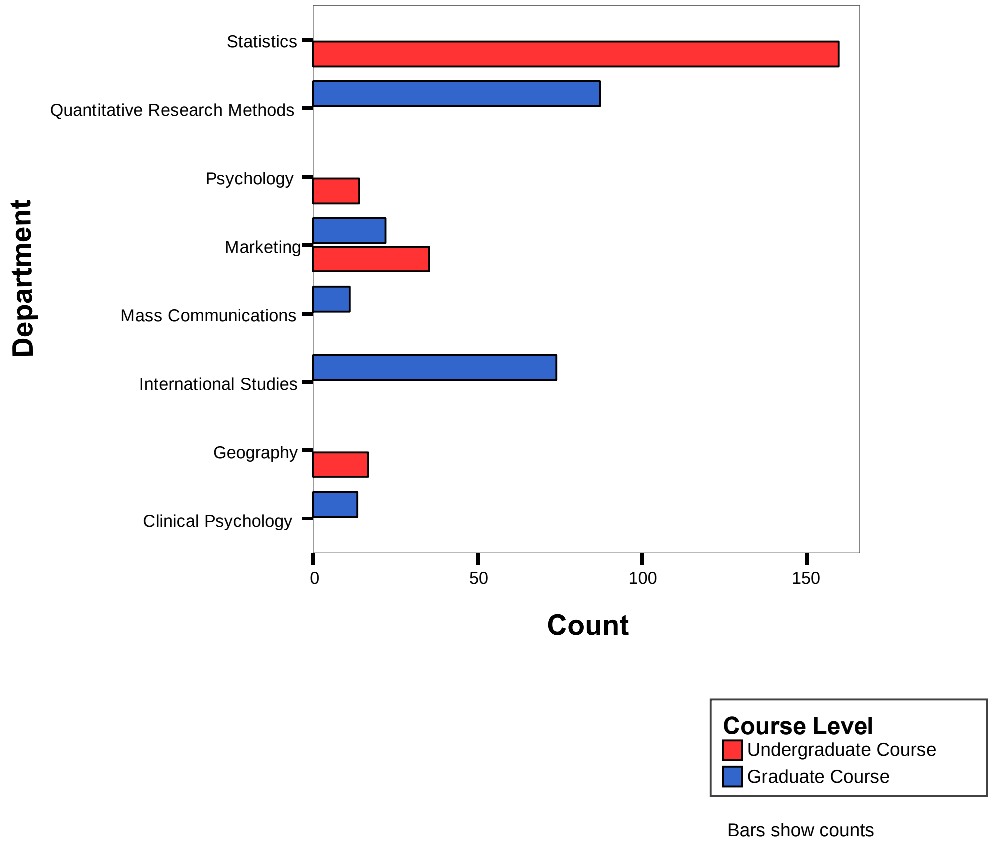

Figure 4. Final Administration Sample Distribution by Department.

Table 4

Final Administration Sample Distribution by Race/Ethnicity

Race

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid		1	.2	.2	.2
	African American	10	2.3	2.3	2.5
	Latin American	21	4.8	4.8	7.4
	Ehananian	1	.2	.2	7.6
	Asian/Pacific Islander	37	8.5	8.5	16.2
	Middle Eastern	3	.7	.7	16.9
	Native American	10	2.3	2.3	19.2
	Caucasian/White	350	80.8	80.8	100.0
	Total	433	100.0	100.0	

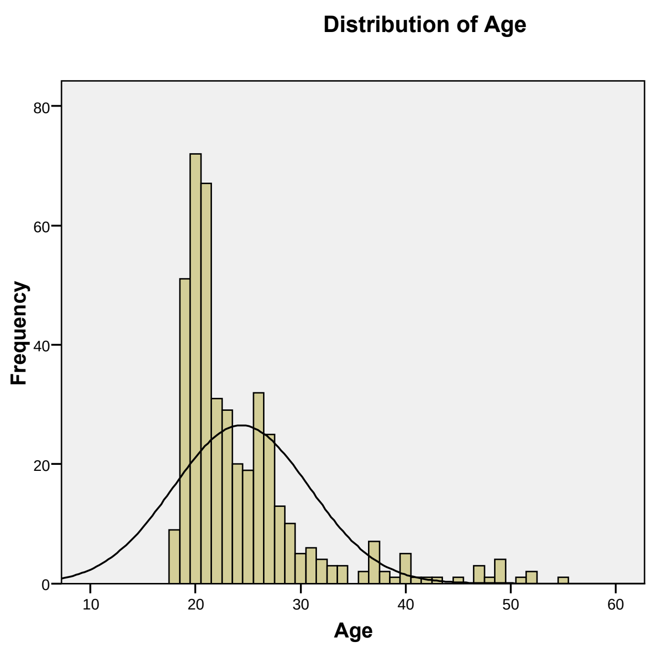

Figure 5. Final Administration Sample Distribution by Age.

Measures

Three measures were used to test the above research questions: the Statistics Anxiety Measure (SAM), the Survey of Attitudes Toward Statistics (SATS: Schau et al., 1995) and the Statistical Anxiety Rating Scale (STARS: Cruise & Wilkins, 1980). The SATS and STARS were chosen due to their demonstrated reliability as multidimensional measures, measuring four or more dimensions of attitude towards statistics and statistics anxiety. As discussed in the preceding chapter, the SATS and STARS were validated using Wiseôs Attitudes Towards Statistics (ATS) measure and were found to significantly correlate with the ATS measure across all dimensions for SATS and all but one for STARS (Mji & Onwuegbuzie, 1987; Schau et al., 1995).

The SAM had 60 four-point rating scale items in the pilot study (10 items belonging to each of the six sub-scales: anxiety, fearful behavior, attitude, expectation, history and self-concept, and performance) and 43 four-point rating scale items distributed across five reformed dimensions based on the results of the pilot study in the field administration (12 items belonging to anxiety, 9 items belonging to attitude towards the class, 4 items belonging to fearful behavior, 10 items belonging to attitude toward math, and 8 items belonging to performance). The SAM assessed a multidimensional construct of statistics anxiety. For the purposes of this study the researcher administered six subscales of the SAM in the pilot study and five subscales in the final administration. No administration exceeded the proposed limit of 30 minutes.

The SATS (Schau et al., 1995) has 28 seven-point rating scaled items and is generally used for assessing attitudes toward statistics (depending on factor and sample, Cronbachôs alpha ranged from .64 to .85, Schau et al., 1995). The SATS is comprised of four subscales: Affect, Cognitive Competence, Value, and Difficulty. The four subscales of the SATS were significantly correlated with the Attitudes Toward Statistics (ATS: Wise, 1985). Course scale and all but Difficulty were correlated with the ATS Field scale (Schau et al.). For validation purposes, the researcher administered the SATS (Schau et al.) in its entirety, and expected administration to take no longer than five minutes.

The STARS (Cruise & Wilkins, 1980) has 51 five-point Likert-scaled items and is generally used for assessing statistics anxiety (Cronbachôs alpha ranged from .68 to .94). All but one of the subscales (Interpretation Anxiety) of the STARS were significantly correlated with the ATS (Mji & Onwuegbuzie, 2004;Wise, 1985). Course scale and all subscales of the STARS were significantly correlated with the ATS Field Scale. The STARS is comprised of six subscales, but only two are directly related to anxiety: ñInterpretation Anxietyò and ñ Test and Class Anxietyò, while the other four measure ñWorth of Statisticsò, ñConceptual Self-Conceptò, ñFear of Asking for Helpò, and ñFear of Statistics Teachersò. For validation purposes, the researcher administered all six subscales of the STARS, and expected administration to take no longer than five minutes.

The SAM, the SATS (Schau et al., 1995), and the STARS (Cruise & Wilkins, 1980) all measure traits using a sum of the raw scores. The researcher based correlational analyses on the raw scores.

Procedure

Pilot. Using a sample of twenty statistics courses being offered in the Winter Quarter 2007, the researcher or the researcherô s assistant distributed the SAM and the SATS in person in two classes (n = 32); and the SAM and the STARS (Cruise & Wilkins, 1980) in another two classes (n = 27); and all three measures in two more classes (n = 48). All classes where cross-validation measures were distributed along with the SAM had a minimum of seven students; the remaining fourteen classes were given only the SAM.

Field Administration. Using a sample of twenty-one statistics classes offered in Spring Quarter 2007, the researcher redistributed the SAM in order to determine if item elimination and possible reconstruction after pilot administration of domains improved the overall fit of the model. The researcher redistributed the STARS and the SAM in one class (n = 7), the SATS and the SAM in two classes (n = 31), and all three measures in an additional three classes (n = 85). One class which was meant to receive the STARS and the SAM accidentally received all three measures, which is why the STARS and the SAM sample was smaller than expected and the STARS, SATS, and SAM sample was greater than expected. The remaining fifteen classes were only given the SAM.

Survey Administration

Pilot. Students were asked to complete the SAM, as well as any cross-validation measures, at the start of class on the first day ideally or the second day depending on the instructorôs schedule and/or preference. Once permission was obtained from the instructor, the researcher or the researcherôs assistant distributed the measure in each class, in person, with a cover sheet that clearly explained that participation was completely voluntary and that the participant was volunteering to participate by completing the given questionnaires (Appendix B). The cover letter contained contact information for the Institutional Review Board, the researcher, as well as the website for gathering further information on the study. Administrations of the measures were planned to occur in the time ranges of 9:00 A.M. to 7:30 P.M January 3-9, 2007. The researcher compensated participants with assorted candy and organic raisins. All the data collected were anonymous. Surveys were placed by participants, the researcher, or the researcherôs assistant in the slot of a locked secure box. In classrooms where more than one measure was administered, validation surveys were folded with the SAM and later coded so that they could be linked when necessary. All data were considered anonymous.

Field Administration. Just as in the pilot study, once permission was obtained from the instructor, the researcher distributed the revised SAM attaching a cover sheet that clearly explained that participation was completely voluntary and that the participant was volunteering to participate by completing the given questionnaires. The cover letters were the same as those used in the pilot study, including contact information for the Institutional Review Board, the researcher, as well as the website for gathering further information on the study. Administrations of the measures occurred in the time ranges of 9:00 A.M. to 9:00 P.M March 27- April 5, 2007. The researcher compensated participants with assorted candy. All the data collected were anonymous. Validation surveys were folded with the SAM and later coded so that they could be linked when necessary. Surveys were placed by participants, the researcher, or the researcher's assistant in the slot of a locked secure box just like students did in the pilot study. All data were considered anonymous.

Analyses

Pilot. Pilot data were analyzed using exploratory factor analysis (EFA) and multidimensional item response theory (MIRT). Factor structure and inter-item correlations were assessed in EFA for guidance in item and factor elimination. Item fit was considered in MIRT, but was mainly used to further guide the researcher in ordering items for the final administration of the SAM based on new factor structures determined by the EFA. Reliability was assessed for the overall construct of SAM as well as each of the six domains using EFA. Item loadings were tested for strength using EFA and item fit was considered in MIRT. Items with loadings below .30, items that cross loaded, and items that demonstrated poor fit were removed before final field administration of the SAM took place. The number of factors was reduced by one, items were reallocated across newly defined factors, and items were reordered within factors based on agreeability findings from the pilot study.

Field Administration. Analysis of the final field administration was done primarily to assess structural reliability. The structure of the SAM was tested using Confirmatory Factor Analysis (CFA) in LISREL. Since the data assumed an ordinal scale, the researcher used PRELIS to calculate asymptotic covariances and polychloric correlations of all four-point rating scale items. Using LISREL weighted least squares estimation the theorized model of the SAM was tested for factor structure. Item fit was reassessed using multidimensional item response theory. The reliability of subscales was assessed using the final field administration data. It was expected that if the SAM was reliable, the remaining items would have good structural fit using the final administration data based on the loadings found using the pilot data.

Phase 4: Validation

Validation

Convergent validity was assessed through two correlational studies. The researcher expected the pilot and final administration anxiety subscales of the SAM to correlate highly with the STARS. The researcher originally expected the pilot and final administration attitude subscales of the SAM to correlate highly with the SATS. The researcher expected the other domains of SAM to be moderately correlated with the STARS and the SATS. Expected correlations are provided in Table 5.

Table 5

Expected Correlations between the Original SAM Domains, STARS, and SATS

SAM Pilot Domains	STARS	SATS
Anxiety	.75-.90	<.40
Performance	<.40	<.40
History and Self-Concept	<.40	<.40
Expectations	<.40	<.40
Attitude	<.40	.75-.90
Fearful Behavior	<.40	<.40

However, after the pilot administration the model was respecified to reflect factors identified by the EFA, resulting in the hypothesized correlations laid out in Table 6.

Table 6

Expected Correlations between the EFA SAM Domains, STARS, and SATS

SAM Pilot Domains	STARS	SATS
Anxiety	.75-.90	<.40
Attitude Towards Class	<.40	.75-90
Fearful Behavior	<.40	<.40
Attitude Towards Math	<.40	.75.90
Performance	<.40	<.40

After the final administration, the model was respecified to reflect factors identified by CFA using five items per factor.

CHAPTER 3

RESULTS

In this chapter, the research questions described in Chapter 1 and the results of scale development phases described in Chapter 2 are addressed. Since the results of phase one, the planning phase, were already discussed at length in Chapter 2, this chapter begins by discussing the results from phase two: construction.

Phase 2: Construction

In phase two: construction, the researcher created an item pool, carried out an expert review of the SAM item pool, selected pilot items based on results of the expert review, and finally conducted two cognitive interviews for item clarification.

Item Pool Creation

The researcher identified six domains of statistics anxiety through thematic analysis by linking visual diagrams of various theories related to statistics anxiety and attitudes toward statistics, as depicted in Figure 1. The researcher defined each of the six factors in Chapter 2, and then projected the number of items to be developed for each of the six factors (Table 2). The researcher ultimately generated an item pool of 120 items based on findings from the literature as discussed in Chapter 1, consideration of items that functioned well on the FST, and the thematic analysis carried out and presented in Figure 1. Items were created to be both distinguishable by domain and level of agreeability.

Expert Review

The researcher conducted five expert reviews of all 120 items to assess content and agreeability validity. These five people included three DU professors, one person in charge of assessment at DU, and one cognitive research methodologist from NASS. These individuals were the Program Chair for the Quantitative Research Methods (QRM) Program for the Morgridge College of Education (MCE) at DU, an assistant professor who teaches survey design classes in the QRM Program, an assistant professor in the Graduate School of Social Work who teaches statistics and assessment-based classes, a doctoral student in the MCEøs Higher Education Program who is the Coordinator of the First-Year Experience Program at DU and was previously responsible quantitative goal assessment for Student Life at DU, and the Chief Cognitive Research Methodologist for NASS who is responsible for in-house survey methodology training and serves as an expert consultant in the review of various NASS surveys and studies.

Each of the five expert reviewers was sent an electronic link to an online survey review of the SAM. The review consisted of 120 items which they were asked to rate in three areas: quality of item, appropriateness of item for the intended domain, and the perceived level of agreeability for that item.

Item Selection

Three of five reviews were completed prior to creation of the pilot version and two were completed after the pilot version was created; although this was not ideal, the researcher was forced to move forward with pilot creation and administration due to scheduled data collection. Using the three expert reviews, the researcher sorted items within each of the six factors by quality (descending order), then by appropriateness of item for the intended domain (descending order), and lastly by perceived agreeability for that item (ascending order). The researcher pooled items from those with the greatest perceived domain fit and quality and used perceived agreeability to order items on the pilot version of the SAM. The last two reviews were used as supplemental support and data relevant to validation of pilot item selection.

Experts were asked to rate 120 items. After items were rated, items were sorted on quality and the top ten items within each factor were selected. This process resulted in 60 items with an overall average quality rating of 4.28 and average item quality ratings ranging from 3.4 to 4.8 using a 5 point rating scale where 1 = very poor and 5 = excellent. All subscales averaged item quality ratings greater than 4: anxiety 4.44, fear 4.30, attitude 4.18, expatiations 4.50, history 4.22, and performance 4.29. The overall average domain appropriateness of the 60 pilot items averaged 4.09 with average item domain appropriateness ratings ranging from 2.8 to 4.8 using a 5 point rating scale. All subscales averaged item domain appropriateness ratings greater than 4, except the expectations factor: anxiety 4.30, fear 4.12, attitude 4.02, expectations 3.57, history 4.24, and performance 4.29. All of the above average ratings were done using ratings from all five expert reviewers. Items were not revised based on expert review, but some items were later combined or reworded based on cognitive interviews.

Pilot items were ordered based on the original three expert review ratings of perceived level of agreeability, from easiest to agree with to hardest to agree with. Surprisingly the agreeability ratings of all five expert reviewers did not significantly correlate with the MIRT logit position estimates using the pilot data for all sixty pilot items ($r = 0.031$, $p > .05$), the 43 final administration items ($r = 0.110$, $p > .05$), or the 23 final model items ($r = 0.052$, $p > .05$). Reasons that may account for the lack of relationship between perceived difficulty of items and actual difficulty, may include: experts not being fully aware of what students would actually agree or disagree with; not all experts having a background in IRT, and thus not fully understanding the rating of agreeability; and/or the rating of agreeability being completely subjective, and thus too personal of a question to extrapolate and generalize beyond the individual responding. In retrospect, it appears the ordering of items and questions of agreeability difficulty should have been carried out via cognitive interviews, since students were the actual audience whom would agree or disagree with items on the SAM.

Cognitive Interviews

Cognitive interviews were conducted with two 21 year old male students who were juniors at DU taking introductory level statistics in the business school. The cognitive interview consisted of a review of all 120 items of the pool prior to administration of the pilot version of the SAM. All 120 items were used for the cognitive interviews since the researcher was still hoping to receive more than three of the five expert reviews before finalizing the pilot version of SAM. Both interviews took place in the home of one of the students. The researcher gave a copy of all 120 items in survey form to each participant at separate times and asked him to think aloud as he worked his way through the survey. Although it was not necessary for either student to actually answer the items, both students did. They asked questions when statements were unclear to them, and the researcher responded by asking what they thought the item was saying. If the response indicated that the item was being interpreted the way it was intended to be, the researcher would ask if they had a better suggestion for phrasing the item. If the response indicated that the item was not being interpreted the way it was intended to be, the researcher explained what she meant and asked the interviewee for suggestions as to how to restate the item so that its meaning was clear. Both interviewees commented when items appeared repetitive in relation to other items. The researcher took notes throughout both interviews and flagged items on the survey that needed to be addressed through a clarification or consolidation process. Both students were rewarded with $50 cash as a ñthank youò for taking time out of their last night of winter break to participate in cognitive interviews. After analysis of expert reviews took place, the researcher used the results of the cognitive interviews to reword the 60 items with the highest overall quality ratings that were too vague or unclear. Four items were changed: 1) Originally students were asked to rate the level of anxiety they experienced when ñInterpreting mathematical solutions,ò but it was recommended that this statement be reworded as ñDeveloping conclusions based on mathematical solutionsò since one student felt ñInterpreting mathematical solutionsò was too vague and its rating would depend on how easy or hard the solution was to arrive at; 2) ñI worry extensively about taking this classò was reworded to state ñI am worried about taking statisticsò since one student felt ñextensivelyò was too subjective and was unsure what class ñthis classò pertained to; 3) ñI expect I will be more hesitant to participate in this classò was reworded to state ñI don't see myself participating in this classò since the wording appeared overly complex to one student; and 4) ñTaking this class makes me tenseò was reworded to state ñTaking this class stresses me outò since the meaning of tense was unclear to one student. One item in particular was combined since the meaning was interchangeable to both students: ñSolving equations using the calculatorò and ñSolving equations using the computerò were combined into one item, ñSolving equations using the calculator/computer.ò In some cases it was recommended that items be dropped since interviewees deemed them repetitive. Items that were recommended for elimination in cognitive interviews were actually eliminated when the 60 items with the lowest quality expert review ratings within their respective domain were removed.

Phase 3: Quantitative Evaluation

In this phase, the following research questions were addressed:

> Do statistics anxiety items generated to reflect the six identified domains (anxiety, fearful behavior, attitude, expectations, history and self-concept, and performance) factor appropriately in the six domains?
>
> Does the Statistics Anxiety Measure evidence adequate reliability and validity?
>
> Does multidimensional item response theory provide better assessment of this multidimensional measure than classical test theory?
>
> Does the SAM significantly correlate with the STARS and the SATS, more specifically does the STARS correlate more highly with the anxiety domain of the SAM, and does the SATS correlate more highly with the attitude domains of SAM?

To answer these questions, item loadings, item fit, and item reliability were assessed using a series of analyses from the pilot administration through the field administration. The analyses included:

(1) pilot data factor structure was explored using EFA,

(2) pilot data item difficulty was assessed using MIRT,

(3) pilot data item fit was assessed using MIRT

(4) hybrid rating scores were created for reduction of final administration items,

(5) final administration factor structure and item error variances were assessed using CFA,

(6) final administration item misfit and difficulty were assed using MIRT, and

(7) cross-validation studies were conducted, testing overall correlations across measures as well as individual factors.

Pilot Measure Structure

First, the factor structure of the pilot data was explored using EFA. The researcher used a sample of 347 cases to conduct an EFA, with the original 60 SAM items as indicators. To test the number of factors indicated by the measurement items, the researcher examined a scree plot (Figure 6) of the eigenvalues using principle axis factoring.

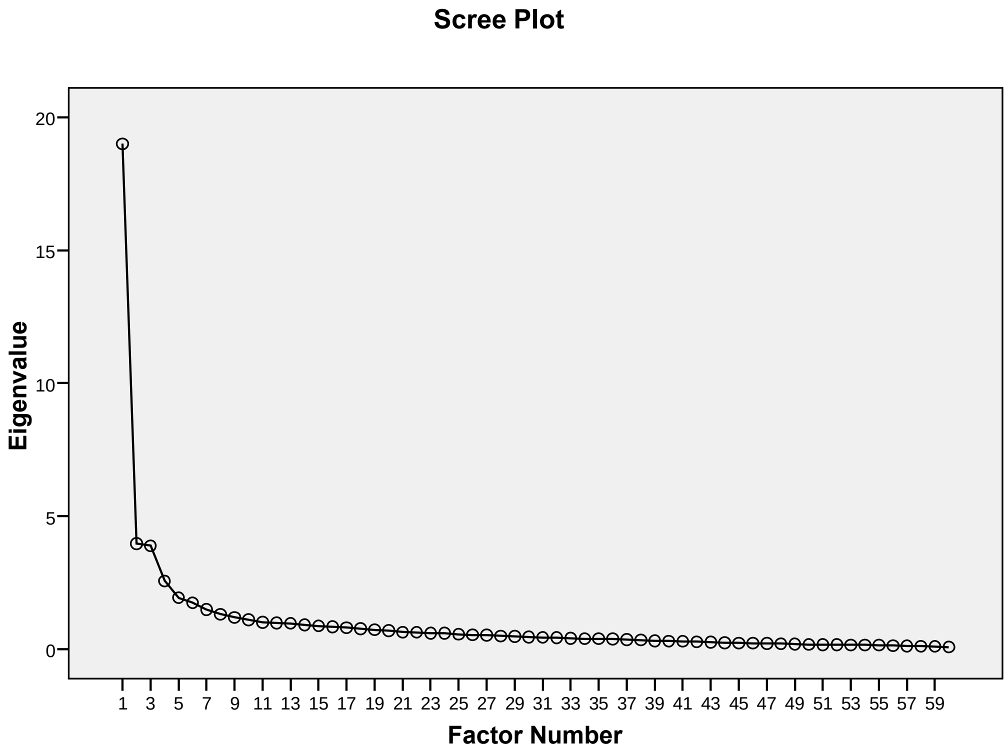

Figure 6. Pilot Data Scree Plot.

The scree plot demonstrated that a four-, five- or six-factor solution fit the data; however, the scree plot did not indicate that a dramatic increase in variance explained occurred by using any more than four-factors. According to Table 7 the total variance explained did not greatly increase when going from four to six factors.

Table 7

Total Variance Explained by Factors in Pilot Data

Total Variance Explained

	Initial Eigenvalues			Extraction Sums of Squared Loadings		
Factor	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	18.995	31.659	31.659	18.574	30.957	30.957
2	3.975	6.624	38.283	3.526	5.876	36.833
3	3.892	6.487	44.770	3.464	5.773	42.606
4	2.565	4.275	49.045	2.071	3.451	46.057
5	1.939	3.232	52.277	1.392	2.320	48.376
6	1.749	2.915	55.192	1.268	2.113	50.489
7	1.500	2.499	57.692			
8	1.317	2.194	59.886			
9	1.200	2.000	61.885			
10	1.103	1.838	63.724			
11	1.010	1.683	65.407			
12	.989	1.648	67.055			
13	.974	1.624	68.678			
14	.915	1.526	70.204			
15	.881	1.468	71.672			
16	.839	1.398	73.070			
17	.815	1.359	74.429			
18	.773	1.288	75.717			
19	.735	1.225	76.942			
20	.697	1.162	78.105			
21	.644	1.073	79.177			
22	.638	1.063	80.240			
23	.610	1.017	81.257			
24	.604	1.006	82.263			
25	.561	.935	83.199			
26	.542	.904	84.102			
27	.535	.891	84.993			
28	.500	.834	85.828			
29	.486	.810	86.638			
30	.461	.768	87.406			
31	.449	.748	88.153			
32	.433	.721	88.874			
33	.415	.692	89.566			
34	.398	.664	90.230			
35	.394	.657	90.887			
36	.385	.641	91.529			
37	.359	.598	92.127			
38	.347	.579	92.706			
39	.319	.532	93.237			
40	.311	.518	93.755			
41	.296	.493	94.248			
42	.282	.469	94.717			
43	.268	.447	95.164			
44	.245	.408	95.572			
45	.235	.392	95.964			
46	.229	.382	96.346			
47	.220	.367	96.713			
48	.208	.347	97.060			
49	.195	.325	97.386			
50	.179	.298	97.683			
51	.171	.285	97.969			
52	.170	.283	98.252			
53	.163	.271	98.522			
54	.160	.267	98.790			
55	.156	.260	99.050			
56	.140	.233	99.283			
57	.121	.202	99.485			
58	.116	.194	99.679			
59	.108	.181	99.860			
60	.084	.140	100.000			

Extraction Method: Principal Axis Factoring.

The researcher chose to examine pattern matrices for both a five- and six-factor structure (Tables 8 and 9), which ultimately demonstrated that factors were easily distinguishable and identifiable (at least 3 items solely loading under each factor at greater than .30) using a five- factor model but not when using a six-factor model. Direct oblimin rotation was used to determine factor loadings since factors were assumed to be correlated. Due to crossloading items, items loading under .30 were eliminated as well as items that continued to crossload even after using a cutoff of .30. This process was repeated until all items loaded under only one factor with loadings greater than .30. After item reduction was complete, the number of items was reduced from 60 to 43. The final 43 items and their loadings are shown in Table 10. After EFA of the final 43 items was complete, items were resorted into a new data file by the EFA factors (Table 11) and were assessed for difficulty in order to determine the order in which they would appear on the final administration version of the SAM (Figure 7). Please note that factor and dimension numbers were changed on Tables 8-11 and Figure *7* to reflect the order in which they are presented in the final administration version of the SAM. Factor/Dimension 1 was color coded using pink, and represents statistics anxiety. Factor/Dimension 2 was color coded using green, and represents attitude towards class. Factor/Dimension 3 was color coded using grey, and represents fearful behavior. Factor/Dimension 4 was color coded using yellow, and represents attitude towards math. Factor/Dimension 5 was color coded using turquoise, and represents performance.

Table 8

Pattern Matrix for Five-factor Solution (All 60 Pilot Items)

Pattern Matrix

	Factor				
Pilot Variable Label	4	1	2	5	3
1. Calculating probabilities	.199	.481	-.088	.107	-.197
2. Taking statistics	.052	.740	.103	.071	.005
3. Developing conclusions based on mathematical solutions	.096	.657	-.056	-.012	-.073
4. Solving mathematical equations	.437	.470	-.083	.109	-.022
5. Reading statistical studies	-.106	.667	.046	-.096	.020
6. Taking a class that involves math	.518	.445	-.037	.079	.035
7. Taking tests in this class	.111	.623	.065	-.044	.060
8. Interpreting statistics	-.112	.820	.006	-.057	.071
9. Explaining your statistical findings	-.102	.718	-.023	-.185	-.034
10. Formulating and testing hypotheses	-.013	.607	-.031	-.135	-.092
1. If there was a way I could avoid taking this class I would	.039	.177	.684	-.050	.147
2. I am worried about taking statistics	.219	.509	.176	-.065	-.130
3. I don't see myself participating in this class	-.061	-.014	.307	-.122	-.399
4. I am worried this class will bring down my GPA	.133	.351	.312	-.166	-.114
5. I worry about doing well in this class	.177	.455	.186	-.038	-.111
6. Taking this class stresses me out	.210	.436	.241	-.016	-.206
7. I've avoided taking this class as long as possible	.125	.063	.204	-.013	-.433
8. I am afraid to ask for help in this class	-.054	.015	.178	.002	-.626
9. I've been worried ever since I was informed this class was a requirement for my degree	.294	.200	.177	.033	-.406
10. I was hesitant to register for this class	.251	.176	.166	.065	-.432
1. I am only taking this class because it is required	.032	.030	.770	-.037	.076
2. I do not expect to enjoy this class	.108	.075	.756	-.036	.079
3. I expect this class to be boring	.048	-.026	.715	-.058	.083
4. Math is my least favorite subject	.846	.013	.112	.012	.071
5. I dislike math	.875	.012	.131	.059	.032
6. I dislike working with numbers	.792	.033	.100	.064	-.049
7. I will never use what I learn in this class again	-.026	-.011	.488	-.092	-.255
8. There is no room to be creative in statistics	.006	.010	.412	-.018	-.202
9. Taking this class will have little impact on my life	-.081	-.042	.683	.053	-.111
10. My ability to calculate statistics will not affect my chances of getting a job in my chosen field	-.109	.066	.640	-.030	-.030
1. Math is the subject where I have the least amount of confidence	.862	.032	-.021	-.024	.056
2. I am not good at math	.809	.111	-.058	-.045	-.036
3. My math reasoning ability is low	.727	.130	-.061	-.108	-.095
4. My major involves working with numbers	-.217	.012	.012	.143	.009
5. My mother or father did well in math related subjects	-.071	.020	.074	.065	.236
6. I lack academic self confidence	.005	.247	-.119	-.047	-.428
7. My parents expect me to do well in this class	-.120	-.021	.099	.260	.229
8. Most people do well in statistics	.040	-.050	.100	.102	-.014
9. I am expected to do well in statistics	-.161	.050	.014	.242	.242
10. I feel helpless when it comes to solving math problems	.475	.184	-.012	-.029	-.232
1. I have never enjoyed classes that involve math	.673	-.031	.211	-.024	-.064
2. I am taking the minimum number of math courses required for my degree	.309	.046	.500	.045	.006
3. I lack motivation to learn or continue learning statistics	.188	.033	.530	-.086	-.122
4. Based on past experience, I expect the material covered in this class and the exams to be difficult	.201	.377	.246	-.060	.015
5. This is my first statistics class ever	.006	.132	.074	.050	-.228
6. My least enjoyable experiences in school involved math	.733	.011	.107	.041	-.071
7. I've never enjoyed working with numbers	.660	.051	.125	.031	-.165
8. I've struggled to follow the material covered in statistics classes in the past	.237	.335	.115	-.174	-.012
9. I doubt my ability to learn statistics in a lecture environment	.199	.245	.281	-.184	-.100
10. I have low self-esteem when it comes to math	.646	.168	-.071	.009	-.249
1. Explaining my answers	-.037	.042	.060	-.451	-.297
2. Quizzes	.180	.025	.168	-.643	.063
3. Projects	-.119	.045	.013	-.382	-.286
4. Solving mathematical equations	.542	-.085	.037	-.499	.092
5. Solving equations by hand	.552	-.031	-.087	-.466	.098
6. Interpreting my answers	-.058	.073	.089	-.626	-.092
7. Making accurate conclusions based on statistical findings	-.081	.169	.116	-.642	.035
8. Exams	.136	.181	.159	-.529	.103
9. Solving equations using the calculator/computer	.176	.017	-.027	-.475	-.148
10. Developing appropriate methodology to test a given hypothesis	-.063	.230	-.020	-.549	-.081

Extraction Method: Principal Axis Factoring.
Rotation Method: Oblimin with Kaiser Normalization.
a Rotation converged in 15 iterations.

Table 9

Pattern Matrix for Six Factor Solution (All 60 Pilot Items)

Pattern Matrix

	Factor					
Pilot Variable Label	4	1	2	5	3	6
1. Calculating probabilities	.210	.469	-.062	.117	-.185	-.090
2. Taking statistics	.042	.627	.072	.048	-.181	.190
3. Developing conclusions based on mathematical solutions	.121	.661	-.016	.003	-.060	-.043
4. Solving mathematical equations	.457	.486	-.040	.122	-.014	-.028
5. Reading statistical studies	-.067	.705	.102	-.074	.068	-.035
6. Taking a class that involves math	.528	.442	-.006	.082	.001	.034
7. Taking tests in this class	.090	.502	.020	-.076	-.146	.232
8. Interpreting statistics	-.068	.868	.072	-.031	.112	-.003
9. Explaining your statistical findings	-.073	.727	.017	-.171	-.010	-.034
10. Formulating and testing hypotheses	.021	.648	.023	-.111	-.017	-.111
1. If there was a way I could avoid taking this class I would	.026	.109	.632	-.076	-.023	.232
2. I am worried about taking statistics	.154	.307	.062	-.121	-.442	.269
3. I don't see myself participating in this class	-.089	-.075	.254	-.136	-.422	-.112
4. I am worried this class will bring down my GPA	.050	.124	.172	-.239	-.466	**.321**
5. I worry about doing well in this class	.103	.238	.058	-.099	-.454	**.311**
6. Taking this class stresses me out	.149	.251	.136	-.065	-.480	.201
7. I've avoided taking this class as long as possible	.087	-.022	.146	-.031	-.494	-.096
8. I am afraid to ask for help in this class	-.078	-.031	.143	.000	-.579	-.264
9. I've been worried ever since I was informed this class was a requirement for my degree	.238	.059	.091	.000	-.569	.009
10. I was hesitant to register for this class	.183	.001	.058	.025	-.645	.041
1. I am only taking this class because it is required	.024	-.006	.729	-.055	-.035	.150
2. I do not expect to enjoy this class	.086	.001	.694	-.066	-.093	.214
3. I expect this class to be boring	.052	-.021	.700	-.065	.039	.088
4. Math is my least favorite subject	.838	.025	.127	.007	.040	.038
5. I dislike math	.870	.031	.151	.057	.016	.010
6. I dislike working with numbers	.810	.095	.151	.082	.024	-.109
7. I will never use what I learn in this class again	-.007	.048	.522	-.070	-.131	-.216
8. There is no room to be creative in statistics	.020	.048	.432	-.003	-.123	-.148
9. Taking this class will have little impact on my life	-.056	.020	.719	.073	-.024	-.115
10. My ability to calculate statistics will not affect my chances of getting a job in my chosen field	-.085	.114	.667	-.017	.023	-.051
1. Math is the subject where I have the least amount of confidence	.823	-.033	-.055	-.054	-.082	.142
2. I am not good at math	.786	.084	-.061	-.059	-.093	.027
3. My math reasoning ability is low	.701	.094	-.071	-.123	-.148	.004
4. My major involves working with numbers	-.225	-.022	-.010	.133	-.045	.069
5. My mother or father did well in math related subjects	-.085	-.034	.038	.043	.101	.228
6. I lack academic self confidence	-.010	.200	-.133	-.050	-.406	-.172
7. My parents expect me to do well in this class	-.136	-.082	.060	.239	.078	.249
8. Most people do well in statistics	.043	-.038	.106	.105	-.003	-.014
9. I am expected to do well in statistics	-.174	-.010	-.021	.222	.095	.249
10. I feel helpless when it comes to solving math problems	.431	.080	-.067	-.059	-.361	.039
1. I have never enjoyed classes that involve math	.684	.022	.248	-.012	.001	-.102
2. I am taking the minimum number of math courses required for my degree	.316	.064	.509	.046	-.006	.016
3. I lack motivation to learn or continue learning statistics	.176	.008	.503	-.096	-.166	.006
4. Based on past experience, I expect the material covered in this class and the exams to be difficult	.137	.183	.132	-.121	-.311	**.347**
5. This is my first statistics class ever	-.030	.033	.014	.026	-.343	.038
6. My least enjoyable experiences in school involved math	.709	-.014	.096	.027	-.129	.020
7. I've never enjoyed working with numbers	.675	.106	.170	.050	-.076	-.166
8. I've struggled to follow the material covered in statistics classes in the past	.194	.203	.042	-.216	-.221	.209
9. I doubt my ability to learn statistics in a lecture environment	.153	.114	.201	-.227	-.296	.166
10. I have low self-esteem when it comes to math	.607	.089	-.106	-.014	-.340	-.009
1. Explaining my answers	-.025	.092	.087	-.435	-.141	-.271
2. Quizzes	.146	-.047	.108	-.680	-.043	.126
3. Projects	-.113	.069	.023	-.370	-.171	-.221
4. Solving mathematical equations	.512	-.116	.005	-.523	.025	.084
5. Solving equations by hand	.528	-.051	-.105	-.484	.053	.063
6. Interpreting my answers	-.032	.152	.134	-.612	.077	-.209
7. Making accurate conclusions based on statistical findings	-.082	.158	.103	-.649	.035	.010
8. Exams	.102	.086	.095	-.571	-.051	.193
9. Solving equations using the calculator/computer	.157	-.007	-.048	-.484	-.139	-.071
10. Developing appropriate methodology to test a given hypothesis	-.069	.203	-.036	-.554	-.081	-.035

Extraction Method: Principal Axis Factoring.
Rotation Method: Oblimin with Kaiser Normalization.
a Rotation converged in 26 iterations.

Table 10

Final Pattern Matrix for Five-factor Solution (All 43 Pilot Items)

Pattern Matrix

	Factor				
Pilot Variable Label	4	1	2	5	3
1. Calculating probabilities	.215	.492	-.087	-.031	-.131
2. Taking statistics	.019	.735	.094	-.113	-.141
3. Developing conclusions based on mathematical solutions	.120	.611	-.056	.027	-.086
5. Reading statistical studies	-.064	.699	.105	.062	.124
7. Taking tests in this class	.132	.574	.064	.049	.040
8. Interpreting statistics	-.093	.878	.026	-.018	.073
9. Explaining your statistical findings	-.092	.744	-.031	.156	-.018
10. Formulating and testing hypotheses	-.002	.656	-.025	.110	-.008
1. If there was a way I could avoid taking this class I would	.005	.085	.652	.019	-.062
2. I am worried about taking statistics	.216	.462	.095	.061	-.251
5. I worry about doing well in this class	.180	.372	.129	.071	-.188
7. I've avoided taking this class as long as possible	.005	.005	.102	.034	-.648
8. I am afraid to ask for help in this class	-.057	-.046	.122	.126	-.522
9. I've been worried ever since I was informed this class was a requirement for my degree	.233	.152	.046	-.015	-.588
10. I was hesitant to register for this class	.126	.069	.031	-.027	-.706
1. I am only taking this class because it is required	.036	.014	.730	-.002	-.019
2. I do not expect to enjoy this class	.104	.024	.759	-.037	-.014
3. I expect this class to be boring	.064	-.057	.738	.013	.066
4. Math is my least favorite subject	.867	-.013	.068	-.019	.033
5. I dislike math	.910	-.020	.098	-.063	.036
6. I dislike working with numbers	.811	.019	.094	-.070	-.033
7. I will never use what I learn in this class again	-.001	-.032	.472	.165	-.135
8. There is no room to be creative in statistics	-.003	.042	.464	.006	-.090
9. Taking this class will have little impact on my life	-.031	.008	.705	-.041	.035
10. My ability to calculate statistics will not affect my chances of getting a job in my chosen field	-.096	.047	.622	.017	-.018
1. Math is the subject where I have the least amount of confidence	.883	.006	-.086	.039	.016
2. I am not good at math	.824	.086	-.089	.080	-.013
3. My math reasoning ability is low	.758	.099	-.099	.162	-.027
1. I have never enjoyed classes that involve math	.712	-.057	.184	.036	-.004
3. I lack motivation to learn or continue learning statistics	.182	.039	.514	.065	-.099
4. Based on past experience, I expect the material covered in this class and the exams to be difficult	.219	.311	.180	.037	-.097
6. My least enjoyable experiences in school involved math	.768	-.016	.039	.010	-.048
7. I've never enjoyed working with numbers	.700	.028	.095	.018	-.088
8. I've struggled to follow the material covered in statistics classes in the past	.269	.308	.049	.157	-.048
10. I have low self-esteem when it comes to math	.651	.150	-.116	.046	-.202
1. Explaining my answers	-.038	-.027	.042	.557	-.165
2. Quizzes	.201	-.084	.113	.689	.080
3. Projects	-.108	.001	-.019	.504	-.178
6. Interpreting my answers	-.042	.062	.042	.677	.022
7. Making accurate conclusions based on statistical findings	-.011	.124	.030	.690	.113
8. Exams	.153	.057	.130	.574	.095
9. Solving equations using the calculator/computer	.178	.029	-.104	.476	-.099
10. Developing appropriate methodology to test a given hypothesis	-.057	.181	-.077	.600	-.051

Extraction Method: Principal Axis Factoring.
Rotation Method: Oblimin with Kaiser Normalization.
a Rotation converged in 13 iterations.

Table 11

Pilot MIRT Item Key

Factor / Dimension	Pilot MIRT Variable Code	Pilot Variable Name	Pilot Variable Label
1	1	P_Anx01	1. Calculating probabilities
1	2	P_Anx02	2. Taking statistics
1	3	P_Anx03	3. Developing conclusions based on mathematical solutions
1	4	P_Anx05	5. Reading statistical studies
1	5	P_Anx07	7. Taking tests in this class
1	6	P_Anx08	8. Interpreting statistics
1	7	P_Anx09	9. Explaining your statistical findings
1	8	P_Anx10	10. Formulating and testing hypotheses
1	9	P_Fear02	2. I am worried about taking statistics
1	10	P_Fear05	5. I worry about doing well in this class
1	11	P_Hist04	4. Based on past experience, I expect the material covered in this class and the exams to be difficult
1	12	P_Hist08	8. I've struggled to follow the material covered in statistics classes in the past
3	13	P_Fear07	7. I've avoided taking this class as long as possible
3	14	P_Fear08	8. I am afraid to ask for help in this class
3	15	P_Fear09	9. I've been worried ever since I was informed this class was a requirement for my degree
3	17	P_Fear01	1. If there was a way I could avoid taking this class I would
2	18	P_Att01	1. I am only taking this class because it is required
2	19	P_Att02	2. I do not expect to enjoy this class
2	20	P_Att03	3. I expect this class to be boring
2	21	P_Att07	7. I will never use what I learn in this class again
2	22	P_Att08	8. There is no room to be creative in statistics
2	23	P_Att09	9. Taking this class will have little impact on my life
2	24	P_Att10	10. My ability to calculate statistics will not affect my chances of getting a job in my chosen field
2	24	P_Fear10	10. I was hesitant to register for this class
2	25	P_Hist03	3. I lack motivation to learn or continue learning statistics
4	26	P_Att04	4. Math is my least favorite subject
4	27	P_Att05	5. I dislike math
4	28	P_Att06	6. I dislike working with numbers
4	29	P_Exp01	1. Math is the subject where I have the least amount of confidence
4	30	P_Exp02	2. I am not good at math
4	31	P_Exp03	3. My math reasoning ability is low
4	32	P_Hist01	1. I have never enjoyed classes that involve math
4	33	P_Hist06	6. My least enjoyable experiences in school involved math
4	34	P_Hist07	7. I've never enjoyed working with numbers
4	35	P_Hist10	10. I have low self-esteem when it comes to math
5	36	P_Perf01	1. Explaining my answers
5	37	P_Perf02	2. Quizzes
5	38	P_Perf03	3. Projects
5	39	P_Perf06	6. Interpreting my answers
5	40	P_Perf07	7. Making accurate conclusions based on statistical findings
5	41	P_Perf08	8. Exams
5	42	P_Perf09	9. Sovling equations using the calculator/computer
5	43	P_Perf10	10. Developing appropriate methodology to test a given hypothesis

MAP OF LATENT DISTRIBUTIONS AND RESPONSE MODEL PARAMETER ESTIMATES

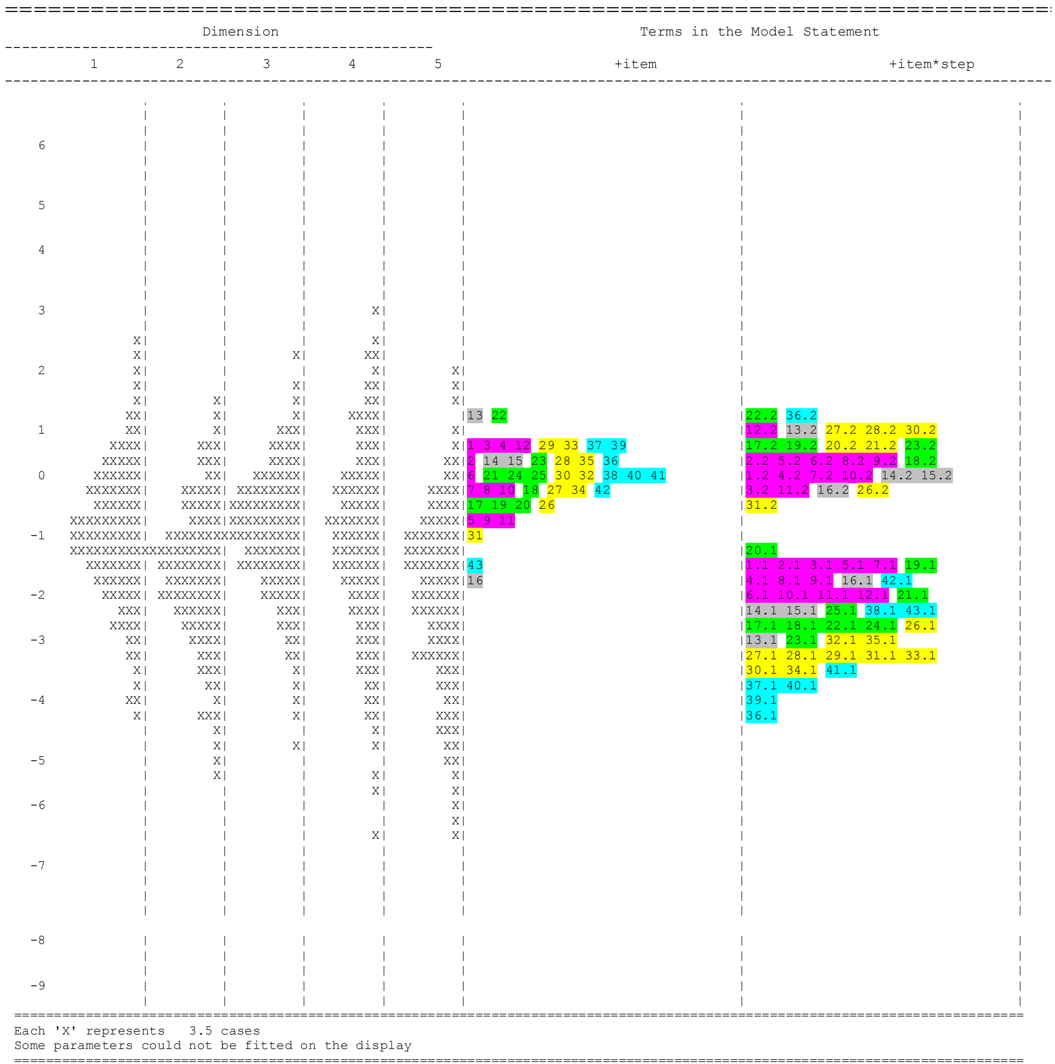

Figure 7. Pilot Data Item Difficulty Plot.[1]

[1] Please refer to Table 11 to relate item numbers used in MIRT to item labels used in EFA.

Final Administration

Determination of Structure Using Confirmatory Factor Analysis

In the final administration study, the researcher used 433 cases to conduct a CFA. Data were imputed for 56 cases (13% of cases) on one or more variables using mode imputation for the given variable where missingness occurred (Table 12). Mode imputation was used in place of mean imputation since data were ordinal.

Table 12

Final Administration Data Missing By Variable

Descriptive Statistics

	N	Missing
7. Taking tests in this class	432	1
9. Explaining your statistical findings	432	1
10. Formulating and testing hypotheses	430	3
8. Interpreting statistics	430	3
2. Taking statistics	431	2
1. Calculating probabilities	429	4
3. Developing conclusions based on mathematical solutions	432	1
5. Reading statistical studies	431	2
2. I am worried about taking statistics	433	0
4. Based on past experience, I expect the material covered in this class and the exams to be difficult	431	2
5. I worry about doing well in this class	433	0
8. I've struggled to follow the material covered in statistics classes in the past	420	13
1. If there was a way I could avoid taking this class I would	433	0
2. I do not expect to enjoy this class	432	1
3. I expect this class to be boring	430	3
1. I am only taking this class because it is required	430	3
7. I will never use what I learn in this class again	433	0
10. My ability to calculate statistics will not affect my chances of getting a job in my chosen field	432	1
3. I lack motivation to learn or continue learning statistics	432	1
9. Taking this class will have little impact on my life	431	2
8. There is no room to be creative in statistics	430	3
10. I was hesitant to register for this class	432	1
8. I am afraid to ask for help in this class	431	2
9. I've been worried ever since I was informed this class was a requirement for my degree	430	3
7. I've avoided taking this class as long as possible	432	1
3. My math reasoning ability is low	430	3
4. Math is my least favorite subject	429	4
5. I dislike math	430	3
7. I've never enjoyed working with numbers	429	4
2. I am not good at math	430	3
1. I have never enjoyed classes that involve math	428	5
6. I dislike working with numbers	429	4
10. I have low self-esteem when it comes to math	427	6
1. Math is the subject where I have the least amount of confidence	429	4
6. My least enjoyable experiences in school involved math	427	6
10. Developing appropriate methodology to test a given hypothesis	425	8
9. Solving equations using the calculator/computer	427	6
3. Projects	423	10
7. Making accurate conclusions based on statistical findings	426	7
8. Exams	426	7
1. Explaining my answers	426	7
2. Quizzes	426	7
6. Interpreting my answers	426	7
Valid N (listwise)	377	56

Prior to conducting a CFA, the number of items had to be reduced due to the limited sample size. In order for a model to be identified in CFA, the sample size must be equal to or exceed the number of distinct sample moments. For 43 items, the required sample size was 946.

In order to reduce the number of items in the Final Administration data, EFA loadings, MIRT item misfit, and reduction in Cronbachῷs alpha reliability estimates were assessed using the pilot data. Three indicator variables were created using the pilot data in order to determine which items should be kept based on EFA loadings, MIRT item misfit, and reduction in Cronbachῷs alpha if item were deleted estimates. In order to create the first indicator, items were first sorted by factor/dimension number (ascending), then MIRT difficulty (ascending), and lastly the absolute value of EFA loading (descending) (Table 13). Of the items with the same factor/dimension number and MIRT difficulty, the one with the highest absolute EFA loading value was coded 1 to indicate that the item should be kept, and all items with unique MIRT difficulty estimates for their respective factor were automatically coded 1.

Table 13

Final Administration Items to Keep Based on EFA Loadings for a 5 Factor Solution Using Pilot Data

Pilot Variable Label	Factor / Dimension	MIRT Difficulty	EFA Loading	EFA Keep Indicator
7. Taking tests in this class	1	5	0.574	1
8. Interpreting statistics	1	7	0.744	1
2. Taking statistics	1	7	0.656	
3. Developing conclusions based on mathematical solutions	1	8	0.878	1
5. Reading statistical studies	1	9	0.735	1
4. Based on past experience, I expect the material covered in this class and the exams to be difficult	1	10	0.492	
5. I worry about doing well in this class	1	10	0.611	
2. I am worried about taking statistics	1	10	0.699	1
10. Formulating and testing hypotheses	1	5	0.462	
9. Explaining your statistical findings	1	5	0.311	
1. Calculating probabilities	1	7	0.372	
8. I've struggled to follow the material covered in statistics classes in the past	1	10	0.308	
1. If there was a way I could avoid taking this class I would	2	6	0.652	
3. I expect this class to be boring	2	6	0.759	1
2. I do not expect to enjoy this class	2	6	0.738	
1. I am only taking this class because it is required	2	7	0.73	1
7. I will never use what I learn in this class again	2	8	0.472	
10. My ability to calculate statistics will not affect my chances of getting a job in my chosen field	2	8	0.622	1
3. I lack motivation to learn or continue learning statistics	2	8	0.514	
9. Taking this class will have little impact on my life	2	9	0.705	1
8. There is no room to be creative in statistics	2	12	0.464	1
10. I was hesitant to register for this class	3	1	-0.706	1
8. I am afraid to ask for help in this class	3	9	-0.522	
9. I've been worried ever since I was informed this class was a requirement for my degree	3	9	-0.588	1
7. I've avoided taking this class as long as possible	3	12	-0.648	1
3. My math reasoning ability is low	4	4	0.758	1
4. Math is my least favorite subject	4	6	0.867	1
7. I've never enjoyed working with numbers	4	7	0.91	1
5. I dislike math	4	7	0.7	
1. I have never enjoyed classes that involve math	4	8	0.824	1
2. I am not good at math	4	8	0.712	
6. I dislike working with numbers	4	9	0.811	1
10. I have low self-esteem when it comes to math	4	9	0.651	
1. Math is the subject where I have the least amount of confidence	4	10	0.883	1
6. My least enjoyable experiences in school involved math	4	10	0.768	
10. Developing appropriate methodology to test a given hypothesis	5	2	0.6	1
9. Sovling equations using the calculator/computer	5	7	0.476	1
3. Projects	5	8	0.504	
7. Making accurate conclusions based on statistical findings	5	8	0.677	1
8. Exams	5	8	0.574	
1. Explaining my answers	5	9	0.557	1
2. Quizzes	5	10	0.689	
6. Interpreting my answers	5	10	0.69	1

In order to create the second indicator, items were first sorted by factor/dimension number (ascending), then MIRT difficulty (ascending), and lastly the absolute value of MIRT Weighted Misfit *t* score (ascending) (Table 14). Of the items with the same factor/dimension number and MIRT difficulty, the one with the lowest absolute MIRT Weighted Misfit *t* score was coded 1 to indicate that the item should be kept, and all items with unique MIRT difficulty estimates for their respective factor were automatically coded 1.

Table 14

Final Administration Items to Keep Based on MIRT Weighted Misfit t Scores for a 5 Factor Solution Using Pilot Data

Pilot Variable Label	Factor / Dimension	MIRT Difficulty	MIRT Item Misfit *t*	Misfit Keep Indicator
7. Taking tests in this class	1	5	4.3	
8. Interpreting statistics	1	7	2.6	
2. Taking statistics	1	7	4.5	
3. Developing conclusions based on mathematical solutions	1	8	4.1	1
5. Reading statistical studies	1	9	1.8	1
4. Based on past experience, I expect the material covered in this class and the exams to be difficult	1	10	2.5	
5. I worry about doing well in this class	1	10	2.7	
2. I am worried about taking statistics	1	10	5.2	
10. Formulating and testing hypotheses	1	5	1.8	1
9. Explaining your statistical findings	1	5	1.9	
1. Calculating probabilities	1	7	-0.8	1
8. I've struggled to follow the material covered in statistics classes in the past	1	10		1
1. If there was a way I could avoid taking this class I would	2	6		1
3. I expect this class to be boring	2	6	0	
2. I do not expect to enjoy this class	2	6	3.8	
1. I am only taking this class because it is required	2	7	-0.2	1
7. I will never use what I learn in this class again	2	8	4.1	
10. My ability to calculate statistics will not affect my chances of getting a job in my chosen field	2	8	0.5	
3. I lack motivation to learn or continue learning statistics	2	8		1
9. Taking this class will have little impact on my life	2	9	-1.4	1
8. There is no room to be creative in statistics	2	12	4.2	1
10. I was hesitant to register for this class	3	1	-3.1	1
8. I am afraid to ask for help in this class	3	9	1.6	1
9. I've been worried ever since I was informed this class was a requirement for my degree	3	9	3.8	
7. I've avoided taking this class as long as possible	3	12	4.6	1
3. My math reasoning ability is low	4	4	2.1	1
4. Math is my least favorite subject	4	6	-5	1
7. I've never enjoyed working with numbers	4	7	-5.4	
5. I dislike math	4	7	1.1	1
1. I have never enjoyed classes that involve math	4	8	3	
2. I am not good at math	4	8	-2.4	1
6. I dislike working with numbers	4	9	-6.4	
10. I have low self-esteem when it comes to math	4	9		1
1. Math is the subject where I have the least amount of confidence	4	10	-0.3	1
6. My least enjoyable experiences in school involved math	4	10	-4	
10. Developing appropriate methodology to test a given hypothesis	5	2		1
9. Sovling equations using the calculator/computer	5	7	-1.2	1
3. Projects	5	8	1	
7. Making accurate conclusions based on statistical findings	5	8	-3.4	
8. Exams	5	8	0	1
1. Explaining my answers	5	9	0.1	1
2. Quizzes	5	10	-0.4	1
6. Interpreting my answers	5	10	-2	

In order to create the third indicator, items were first sorted by factor/dimension number (ascending), then MIRT difficulty (ascending), and lastly estimate of Cronbachôs alpha if the item were deleted (ascending) (Table 15). Of the items with the same factor/dimension number and MIRT difficulty, the one with the lowest estimate of Cronbachôs alpha if item were deleted was coded 1 to indicate that the item should be kept, and all items with unique MIRT difficulty estimates for their respective factor were automatically coded 1.

Table 15

Final Administration Items Recommended to Keep Based on Estimate of Cronbach's Alpha if Item were Deleted for a 5 Factor Solution Using Pilot Data

Pilot Variable Label	Factor / Dimension	Cronbach's α if Item were Deleted	Reliability Keep Indicator
7. Taking tests in this class	1	0.951	
8. Interpreting statistics	1	0.951	1
2. Taking statistics	1	0.951	
3. Developing conclusions based on mathematical solutions	1	0.951	1
5. Reading statistical studies	1	0.951	1
4. Based on past experience, I expect the material covered in this class and the exams to be difficult	1	0.951	1
5. I worry about doing well in this class	1	0.951	
2. I am worried about taking statistics	1	0.952	
10. Formulating and testing hypotheses	1	0.95	1
9. Explaining your statistical findings	1	0.951	
1. Calculating probabilities	1	0.951	
8. I've struggled to follow the material covered in statistics classes in the past	1	0.951	
1. If there was a way I could avoid taking this class I would	2	0.952	
3. I expect this class to be boring	2	0.951	1
2. I do not expect to enjoy this class	2	0.952	
1. I am only taking this class because it is required	2	0.952	1
7. I will never use what I learn in this class again	2	0.952	
10. My ability to calculate statistics will not affect my chances of getting a job in my chosen field	2	0.952	
3. I lack motivation to learn or continue learning statistics	2	0.951	1
9. Taking this class will have little impact on my life	2	0.952	1
8. There is no room to be creative in statistics	2	0.952	1
10. I was hesitant to register for this class	3	0.951	1
8. I am afraid to ask for help in this class	3	0.952	
9. I've been worried ever since I was informed this class was a requirement for my degree	3	0.951	1
7. I've avoided taking this class as long as possible	3	0.952	1
3. My math reasoning ability is low	4	0.95	1
4. Math is my least favorite subject	4	0.951	1
7. I've never enjoyed working with numbers	4	0.95	1
5. I dislike math	4	0.951	
1. I have never enjoyed classes that involve math	4	0.95	1
2. I am not good at math	4	0.951	
6. I dislike working with numbers	4	0.951	
10. I have low self-esteem when it comes to math	4	0.95	1
1. Math is the subject where I have the least amount of confidence	4	0.951	1
6. My least enjoyable experiences in school involved math	4	0.951	
10. Developing appropriate methodology to test a given hypothesis	5	0.952	1
9. Sovling equations using the calculator/computer	5	0.952	1
3. Projects	5	0.952	
7. Making accurate conclusions based on statistical findings	5	0.952	
8. Exams	5	0.951	1
1. Explaining my answers	5	0.952	1
2. Quizzes	5	0.952	1
6. Interpreting my answers	5	0.952	

A fourth indicator was created by calculating a hybrid score variable. The hybrid score variable was calculated by taking the sum of the three previously discussed indicators. The hybrid score ranged from 0-3; 0 meaning it was not recommended by any of the indicators; 1 meaning it was recommended by one indicator; 2 meaning it was recommended by two of the indicators; and 3 meaning it was recommended by all three indicators. In order to create the fourth indicator items were first sorted by factor/dimension number (ascending), then MIRT difficulty (ascending), and lastly the newly created hybrid score (descending) (Table 16).

Table 16

Final Administration Items Recommended to Keep Based on Hybrid Score for a 5 Factor Solution Using Pilot Data

Pilot Variable Label	EFA Keep Indicator	Misfit Keep Indicator	Reliability Keep Indicator	Hybrid Score	Hybrid Keep Indicator
10. Formulating and testing hypotheses		1	1	2	1
8. Interpreting statistics	1		1	2	1
3. Developing conclusions based on mathematical solutions	1	1	1	3	1
5. Reading statistical studies	1	1	1	3	1
4. Based on past experience, I expect the material covered in this class and the exams to be difficult			1	1	1
3. I expect this class to be boring	1		1	2	1
1. I am only taking this class because it is required	1	1	1	3	1
3. I lack motivation to learn or continue learning statistics		1	1	2	1
9. Taking this class will have little impact on my life	1	1	1	3	1
8. There is no room to be creative in statistics	1	1	1	3	1
10. I was hesitant to register for this class	1	1	1	3	1
9. I've been worried ever since I was informed this class was a requirement for my degree	1		1	2	1
7. I've avoided taking this class as long as possible	1	1	1	3	1
3. My math reasoning ability is low	1	1	1	3	1
4. Math is my least favorite subject	1	1	1	3	1
7. I've never enjoyed working with numbers	1		1	2	1
1. I have never enjoyed classes that involve math	1		1	2	1
10. I have low self-esteem when it comes to math		1	1	2	1
1. Math is the subject where I have the least amount of confidence	1	1	1	3	1
10. Developing appropriate methodology to test a given hypothesis	1	1	1	3	1
9. Sovling equations using the calculator/computer	1	1	1	3	1
8. Exams		1	1	2	1
1. Explaining my answers	1	1	1	3	1
2. Quizzes		1	1	2	1
7. Taking tests in this class	1			1	
9. Explaining your statistical findings				0	
2. Taking statistics				0	
1. Calculating probabilities		1		1	
5. I worry about doing well in this class				0	
2. I am worried about taking statistics	1			1	
8. I've struggled to follow the material covered in statistics classes in the past		1		1	
1. If there was a way I could avoid taking this class I would		1		1	
2. I do not expect to enjoy this class				0	
7. I will never use what I learn in this class again				0	
10. My ability to calculate statistics will not affect my chances of getting a job in my chosen field	1			1	
8. I am afraid to ask for help in this class		1		1	
5. I dislike math		1		1	
2. I am not good at math		1		1	
6. I dislike working with numbers	1			1	
6. My least enjoyable experiences in school involved math				0	
3. Projects				0	
7. Making accurate conclusions based on statistical findings	1			1	
6. Interpreting my answers	1			1	

Of the items with the same factor/dimension number and MIRT difficulty, the one with the highest hybrid score was coded 1 to indicate that the item should be kept, and all items with unique MIRT difficulty estimates for their respective factor were automatically coded 1. Ultimately the decision to keep or drop items was based on the hybrid score; items with a hybrid keep indicator of one were kept, and all others were dropped.

The data collected using the SAM are considered ordinal, based on use of a 4-point rating scale. When data are considered ordinal, Jöreskog and Sörbom (1993) recommend using PRELIS to calculate the asymptotic covariances and polychloric correlations of all items modeled, and LISREL with weighted least squares estimation to test the structure of the data. These recommendations were followed to test the structure of a one-, four- and five-factor model using CFA. Analysis of items in the above table resulted in the following five-factor model (Figure 8). However, in order to be able to obtain adequate subscale reliability the researcher chose to collapse the statistics anxiety and fearful behavior factors/dimensions since they were highly correlated with one another to begin with and intuitively represented very similar traits (Table 17). Note that due to negative error variances in the model, items are not perfectly correlated with themselves as one would expect.

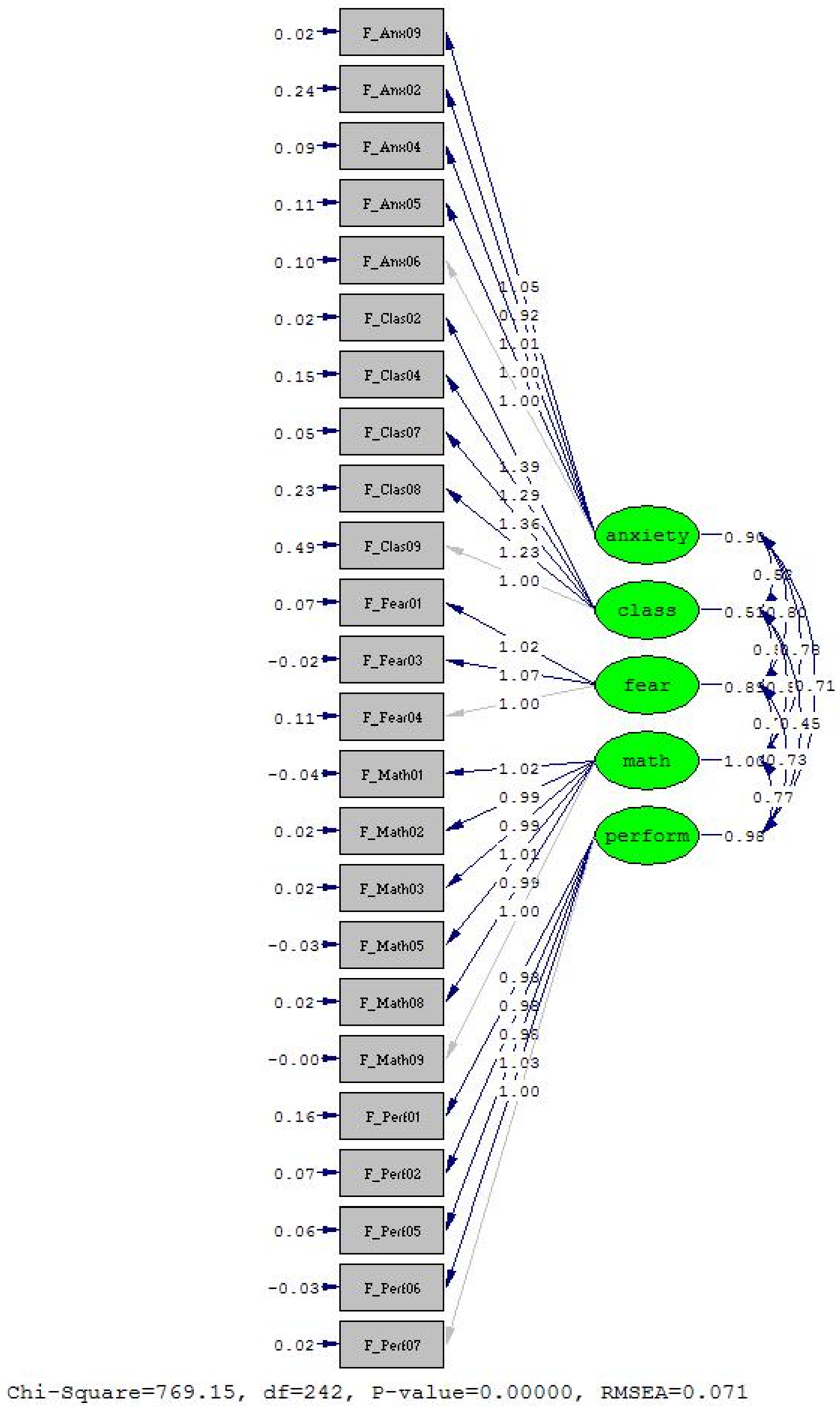

Figure 8. Confirmatory Factor Analysis 5 Factor Solution Path Diagram of R Estimates Based on Hybrid Keep Indicator Using Final Administration Data

Table 17

Confirmatory Factor Analysis Five-factor Solution Correlations and Covariances Using Final Administration Data

	anxiety	class	fear	math	perform
anxiety	0.90 (0.02) 43.41				
class	0.52 (0.02) 25.45	0.51 (0.03) 16.94			
fear	0.80 (0.02) 37.55	0.55 (0.02) 23.83	0.89 (0.03) 31.83		
math	0.78 (0.02) 44.90	0.53 (0.02) 23.87	0.79 (0.03) 30.78	1.00 (0.01) 115.59	
perform	0.71 (0.02) 37.92	0.45 (0.02) 21.58	0.73 (0.02) 32.73	0.77 (0.02) 42.02	0.98 (0.02) 42.65

Correlation
(Significance)
Covariance

Collapsing the statistics anxiety and fearful behavior factors resulted in two items representing the same level of difficulty within the same factor; therefore, the item with the lowest hybrid score was dropped from the following four-factor solution table (Table 18) resulting in the following four-factor model (Figure 9).

Table 18

Final Administration Items Recommended for Keeping Based on Hybrid Score for a 4 Factor Solution Using Pilot Data

Pilot Variable Label	EFA Keep Indicator	Misfit Keep Indicator	Reliability Keep Indicator	Hybrid Score	Hybrid Keep Indicator
10. I was hesitant to register for this class	1	1	1	3	1
10. Formulating and testing hypotheses		1	1	2	1
7. Taking tests in this class	1			1	
9. Explaining your statistical findings				0	
8. Interpreting statistics	1		1	2	1
2. Taking statistics				0	
1. Calculating probabilities		1		1	
3. Developing conclusions based on mathematical solutions	1	1	1	3	1
5. Reading statistical studies	1	1	1	3	1
9. I've been worried ever since I was informed this class was a requirement for my degree	1		1	2	1
8. I am afraid to ask for help in this class		1		1	
4. Based on past experience, I expect the material covered in this class and the exams to be difficult			1	1	1
5. I worry about doing well in this class				0	
2. I am worried about taking statistics	1			1	
8. I've struggled to follow the material covered in statistics classes in the past		1		1	
7. I've avoided taking this class as long as possible	1	1	1	3	1
3. I expect this class to be boring	1		1	2	1
1. If there was a way I could avoid taking this class I would		1		1	
2. I do not expect to enjoy this class				0	
1. I am only taking this class because it is required	1	1	1	3	1
3. I lack motivation to learn or continue learning statistics		1	1	2	1
7. I will never use what I learn in this class again				0	
10. My ability to calculate statistics will not affect my chances of getting a job in my chosen field	1			1	
9. Taking this class will have little impact on my life	1	1	1	3	1
8. There is no room to be creative in statistics	1	1	1	3	1
3. My math reasoning ability is low	1	1	1	3	1
4. Math is my least favorite subject	1	1	1	3	1
7. I've never enjoyed working with numbers	1		1	2	1
5. I dislike math		1		1	
1. I have never enjoyed classes that involve math	1		1	2	1
2. I am not good at math		1		1	
10. I have low self-esteem when it comes to math		1	1	2	1
6. I dislike working with numbers	1			1	
1. Math is the subject where I have the least amount of confidence	1	1	1	3	1
6. My least enjoyable experiences in school involved math				0	
10. Developing appropriate methodology to test a given hypothesis	1	1	1	3	1
9. Sovling equations using the calculator/computer	1	1	1	3	1
8. Exams		1	1	2	1
3. Projects				0	
7. Making accurate conclusions based on statistical findings	1			1	
1. Explaining my answers	1	1	1	3	1
2. Quizzes		1	1	2	1
6. Interpreting my answers	1			1	

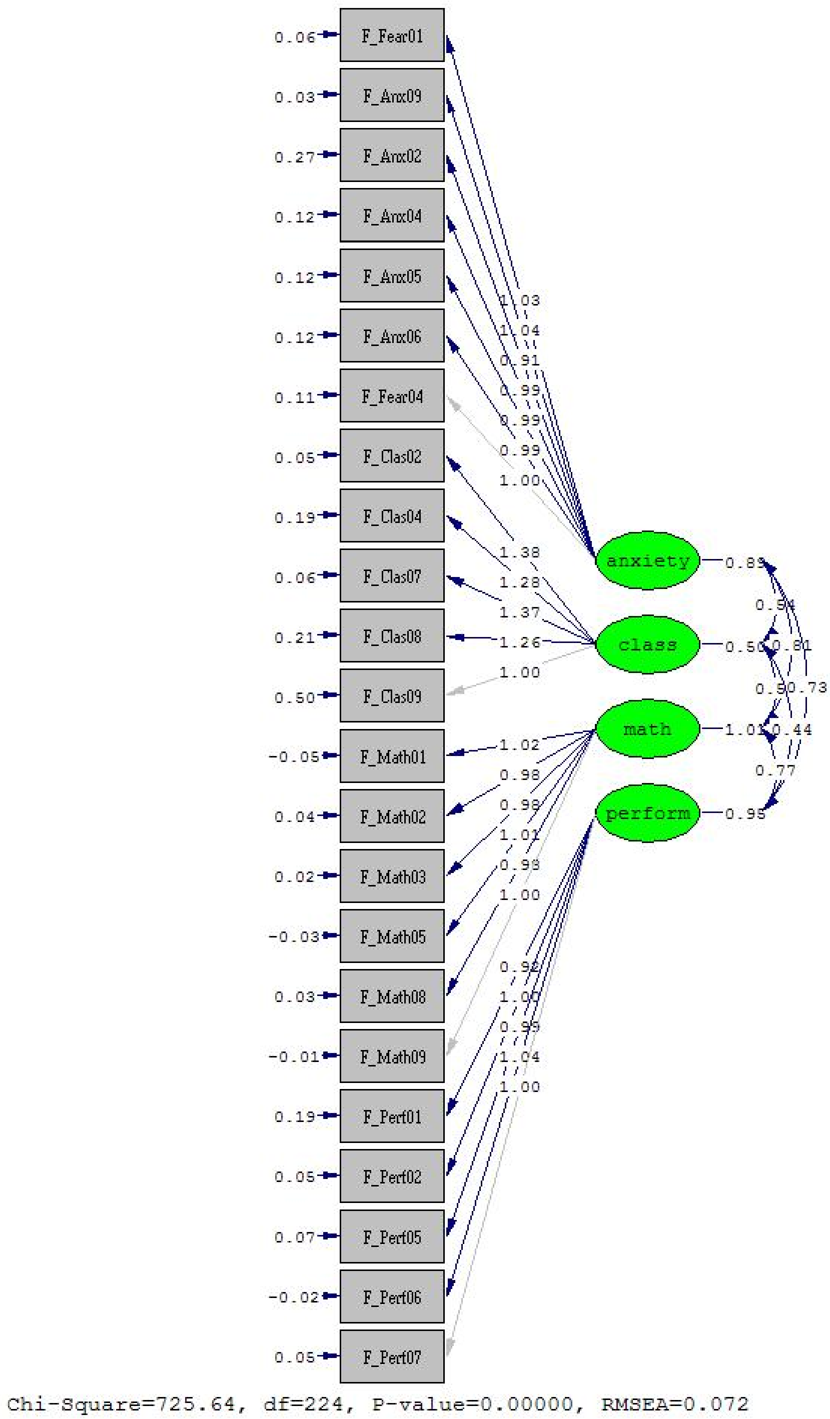

Figure 9. Confirmatory Factor Analysis 4 Factor Solution Path Diagram of R Estimates Based on Hybrid Keep Indicator Using Final Administration Data

Both CFA models were identified (more than two indicators per factor and all factors had a scaling variable). According to Figure 10 and Figure 11, factor loadings in both the five-factor and four-factor models were significant.

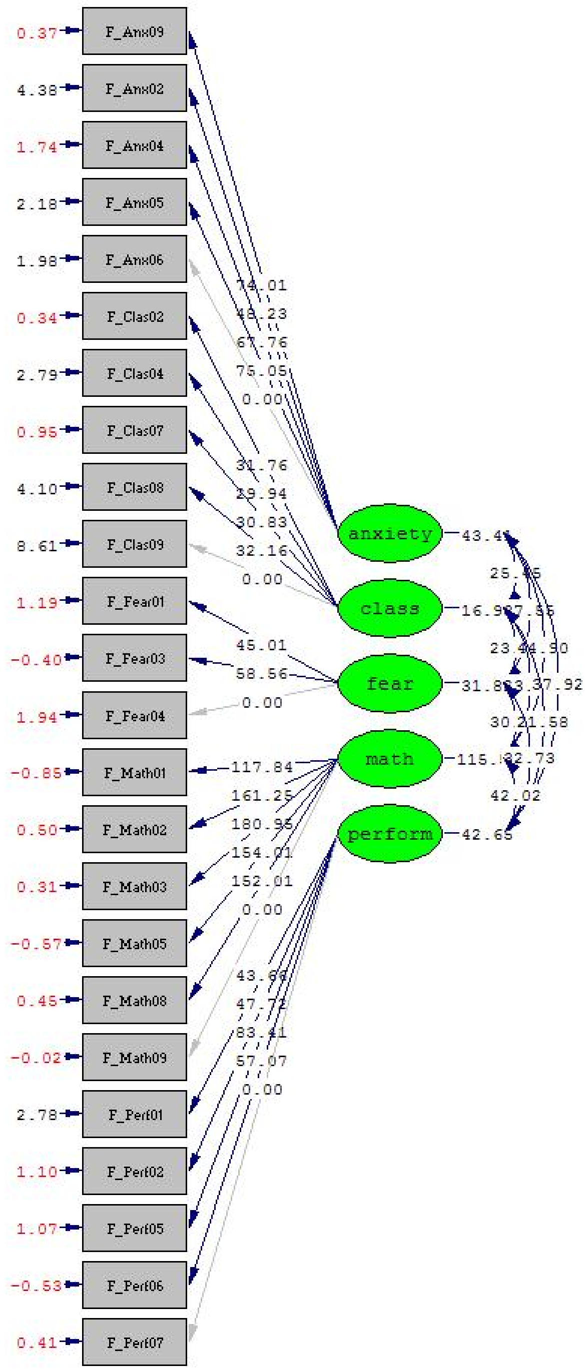

Figure 10. Confirmatory Factor Analysis 5 Factor Solution Path Diagram of *t* Values Based on Hybrid Keep Indicator using Final Administration Data

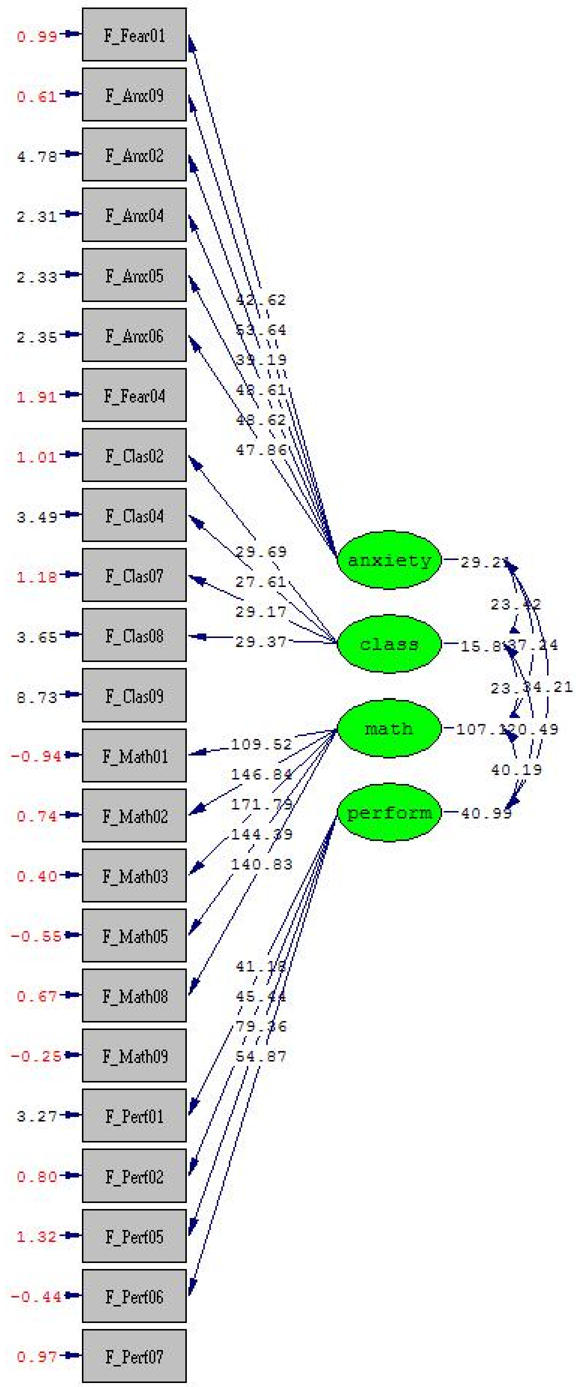

Figure 11. Confirmatory Factor Analysis 4 Factor Solution Path Diagram of *t* Values Based on Hybrid Keep Indicator using Final Administration Data

Using Klineõs (1998) $X^2{}_{difference}$ statistic, the collapsed model using a four-factor solution resulted in a significant increase in model misfit ($X^2{}_{4FactorMeasurementModel} = 725.64$, df = 224, $X^2/df = 3.23$) from the five-factor solution ($X^2{}_{5FactorMeasurementModel} =$ 769.15, df = 242, $X^2/df = 3.17$) ($X^2{}_{difference} =$ 38.99, df = 18, $p < .05$). According to Kline, a favorable index of good fit for large samples is $X^2/df < 3$; however, with smaller sample sizes, Kline points out that even a $X^2/df < 2.5$ could still result from a model with poor overall fit. He recommends turning attention to the ñé indexes that indicate absolute or relative proportions of the observed covariances explained by the model such as the Jöreskog-Sörbom GFI, the Bentler-Bonnett NFI, and the Bentler CFIé ò or ñTheir counterparts that are corrected for the number of parameters (e.g., the AGFI and the NNFI)é ò and recommends that a favorable model will exceed .90 using the above indexes (Kline, 1998, p. 131). Both the five and four-factor models returned NFI, CFI, AGFI, and NNFI values of .99, indicating that 99% of the observed covariance was explained by both models. The Root Mean Square Error of Approximation (RMSEA, Steiger & Lind, 1980) indicated that both models had fair fit according to Steigerõs (1989) and Browne and Cudeckõs (1993) guidelines (RMSEA 0.05-0.08) with minimal difference between the four and five-factor models: $RMSEA_{4FactorMeasurementModel}$ = .072 and $RMSEA_{5FactorMeasurementModel}$ = .071).

To justify a multidimensional construct, the four-factor model was compared with a unidimensional model (Figure 12).

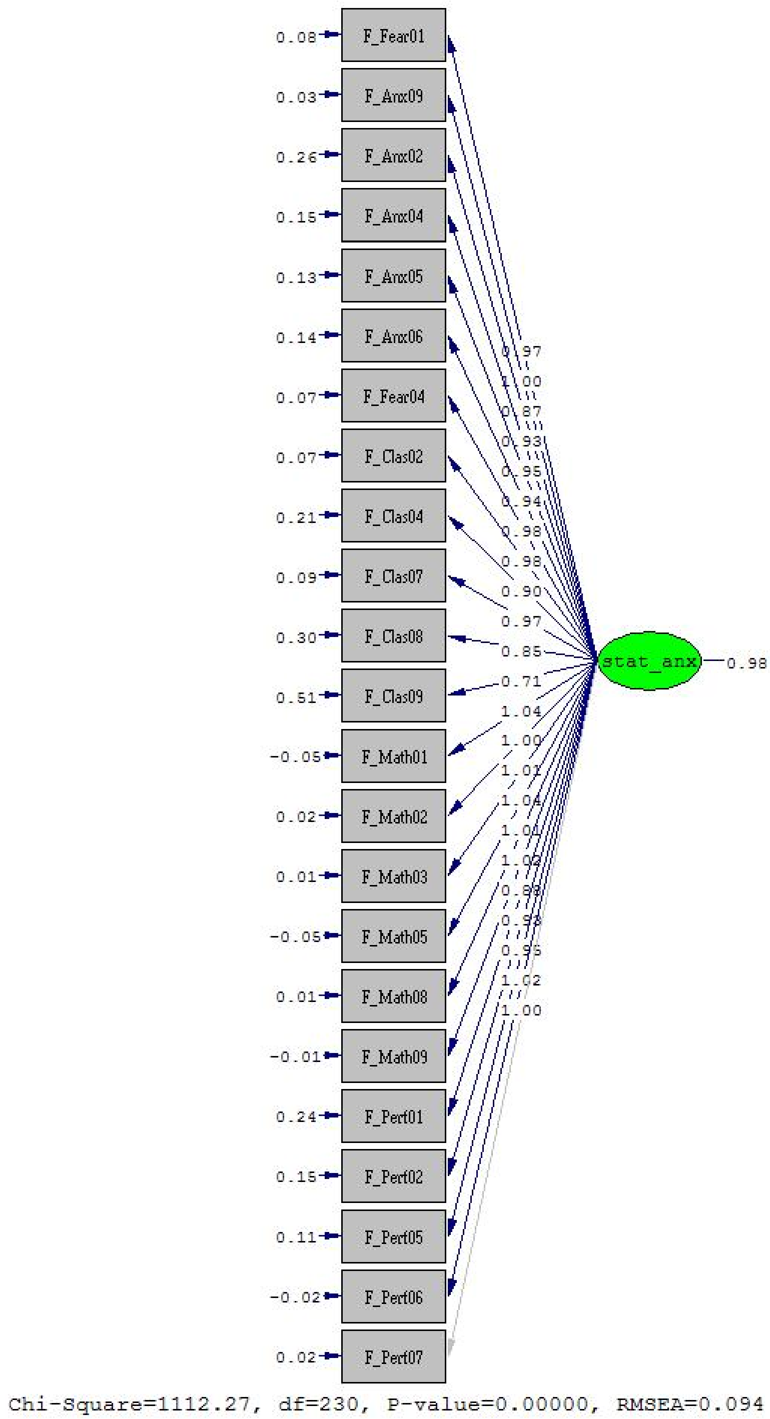

Figure 12. Confirmatory Factor Analysis 1 Factor Solution Path Diagram of R Estimates Based on Hybrid Keep Indicator using Final Administration Data

The one-factor model had significantly greater misfit than the four-factor model [($X^2_{1FactorMeasurementModel} = 1112.27$, *df* = 230, $X^2/df = 4.84$) > ($X^2_{4FactorMeasurementModel} = 725.64$, *df* = 224, $X^2/df = 3.23$), ($X^2_{difference} =$ 386.63, *df* = 6, $p < .05$)]. According to Browne and Cudeckôs (1993) guidelines for interpretation of RMSEA, the RMSEA for the unidimensional model ($RMSEA_{1FactorMeasurementModel}$ = 0.094) had mediocre fit (RMSEA 0.08-0.10); therefore, performing worse than the four-factor model ($RMSEA_{4FactorMeasurementModel}$ = .072) which had fair fit (RMSEA 0.05-0.08).

Since all factor loadings were significant and reasonable fit was achieved with the four-factor model ($X^2_{4FactorMeasurementModel} = 725.64$, *df* = 224, $X^2/df = 3.23$), the four-factor CFA model was tested as a hybrid model with the following paths: Factor 1 (anxiety/fearful behavior), Factor 2 (attitude towards class), Factor 3 (attitude towards math), and Factor 4 (performance). Using Klineôs (1998) two step rule, the measurement model and the structural model were identified; all four-factors were scaled, had more than two indicators, and did not have correlated error variances. Furthermore, the number of observations exceeded the number of specified paths in this model. Referring to Figure 9 and Figure 13, the amount of misfit significantly increased in the four-factor hybrid model compared to the four-factor measurement model [($X^2_{4FactorMeasurementModel} = 725.64$, *df* = 224, $X^2/df = 3.23$) < ($X^2_{HybridModel} = 747.22$, *df* = 226, $X^2/df = 3.31$).), ($X^2_{difference} =$ 10.79, *df* = 2, $p < .05$)].

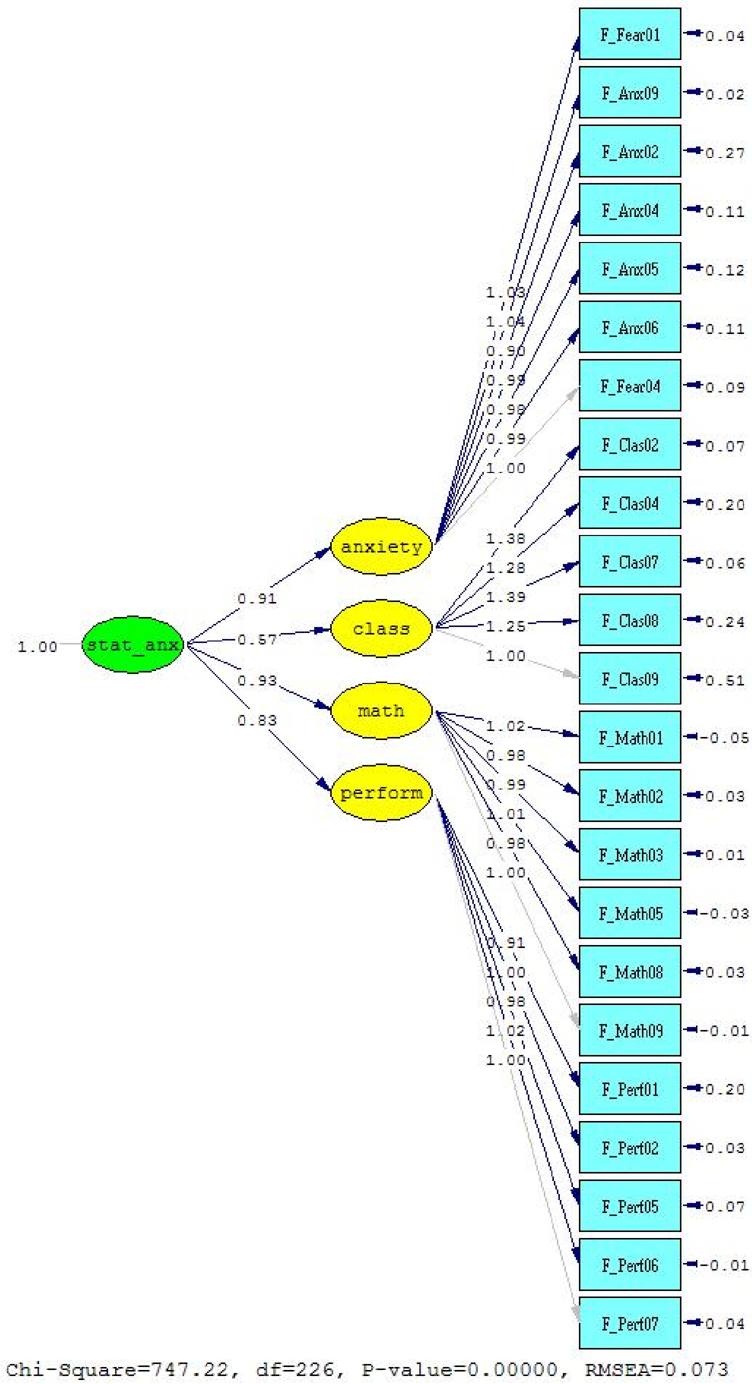

Figure 13. Confirmatory Factor Analysis Hybrid Model Path Diagram of R Estimates

According to Figure 14, the structural portion of the hybrid model was also identified.

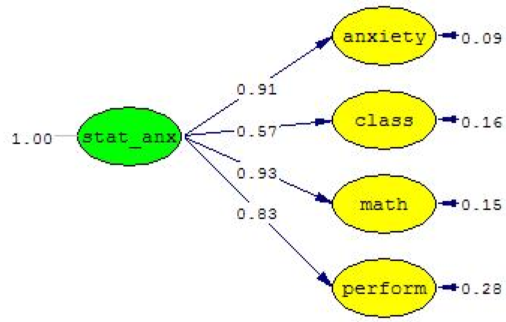

Figure 14. Confirmatory Factor Analysis for the Hybrid Structural Model: Path Diagram of R Estimates

The five-factor model demonstrated significantly less misfit than the unidimensional and the four-factor model; however, the items created to assess fear did not load as expected in the EFA and thus resulted in a factor representing fear that had too few items to adequately measure subscale reliability. Combining fear and anxiety to create a four-factor model where subscale reliability was likely to be higher resulted in significantly more misfit; however, the model still performed significantly better than the unidimensional model. According to Figure 10, the five-factor model produced seven items with significant amounts of misfit ($t > 1.97$), with three items belonging to the anxiety factor, three to the attitude towards class factor, and one from the performance factor. According to Figure 11, the four-factor model resulted in the same seven items having significant amounts of misfit plus one additional item from the anxiety factor ($t > 1.97$). The five-factor, the four-factor, and the unidimensional models all resulted in the same three items having negative error variances, two of which came from the attitude towards math factor and one of which came from the performance factor.

Multidimensional Item Response Theory

The MIRT phase consisted of assessing the four-factor model defined using CFA. Where item fit is assessed through error variances in CFA, item fit is assessed through unweighted (outfit) and weighted (infit) mean square errors in MIRT. According to Bond and Fox (2001), the mean square error (MNSQ) is the mean of the squared residuals for that item, where a residual is calculated by taking ñé the differences between the Rasch modelôs theoretical expectation of item performance and the performance actually encountered for that item in the data matrixò (p. 43). The weighted and unweighted MNSQs differ in that the weighted MNSQs weighs persons performing closer to the item value more heavily; therefore, persons whose ability is more closely matched to the items difficulty level will be weighted more heavily than those who are not (Bond & Fox, 2001). The weighted *t* and the unweighted *t* are just standardized forms of the weighted and unweighted MNSQs, where the MNSQs are transformed to take into account the size of the sample (Bond & Fox, 2001). Since the unweighted MNSQs are more easily influenced by outliers, Bond and Fox (2001) recommend that Rasch modelers pay more attention to the weighted MNSQs. According to Adams and Khoo (1996), items with adequate fit will have weighted MNSQs between .75 and 1.33; however, Bond and Fox state items that are routinely accepted as having adequate fit will have *t* values between -2 and +2. Since the rules are not mutually exclusive; an item may adhere to Adams and Khoo's (1996) standard and still have an absolute value of t that exceeds 2. According to Wilson (2005), when working with large sample sizes, one can expect the t statistic to show significant values for several items regardless of fit; therefore, Wilson suggested that one consider items problematic only if items are identified as misfitting based on *both* the weighted MNSQ and t statistic.

The MIRT model yielded different results from the CFA four-factor model in terms of significant item misfit based on the t statistics ($t < -1.97$ or $t > 1.97$). Using just the t statistic, the CFA four-factor model provided significant error variances for eight items: F_Anx02, F_Anx04, F_Anx05, F_Anx06, F_Class06, F_Class08, F_Class09, and F_Perf01 (Figure 11); where when using Wilson's (2005) rule, the MIRT model resulted in only three items (distinguished in red) demonstrating significant misfit ($t < -1.97$ or $t > 1.97$ and MNSQ < .75 or MNSQ > 1.33) (Table 19).

Table 19

Multidimensional Item Response Theory Response Model Parameter Estimates

TABLES OF RESPONSE MODEL PARAMETER ESTIMATES

TERM 1: item

VARIABLES			UNWGHTED FIT		WGHTED FIT	
item	ESTIMATE	ERROR	MNSQ	T	MNSQ	T
1 F_Fear01	0.541	0.063	0.77	-3.6	0.82	-2.6
2 F_Anx09	-0.586	0.058	0.89	-1.6	0.92	-1.2
3 F_Anx02	-0.495	0.058	1.38	5.1	1.35	4.5
4 F_Anx04	-0.032	0.059	1.27	3.6	1.26	3.5
5 F_Anx05	0.050	0.057	1.06	0.8	1.06	0.8
6 F_Anx06	-0.046	0.057	1.19	2.7	1.22	3.0
7 F_Fear04	0.568*					
8 F_Class02	0.059	0.063	1.17	2.4	1.17	2.4
9 F_Class04	-0.976	0.060	1.34	4.6	1.32	4.3
10 F_Class07	0.378	0.062	1.04	0.6	1.05	0.7
11 F_Class08	0.287	0.064	1.04	0.6	1.06	0.9
12 F_Class09	0.251*					
13 F_Math01	1.080	0.083	1.25	3.4	1.17	2.2
14 F_Math02	-0.802	0.079	0.68	-5.3	0.77	-3.3
15 F_Math03	-0.534	0.079	0.87	-2.0	0.85	-2.1
16 F_Math05	0.415	0.082	0.70	-5.0	0.85	-2.1
17 F_Math08	0.151	0.081	0.60	-6.9	0.68	-4.9
18 F_Math09	-0.310*					
19 F_Perf01	-0.272	0.081	0.91	-1.3	0.95	-0.6
20 F_Perf02	0.510	0.078	1.07	1.0	1.12	1.6
21 F_Perf05	-0.171	0.076	0.39	-11.7	0.41	-10.6
22 F_Perf06	0.205	0.076	0.89	-1.7	0.92	-1.2
23 F_Perf07	-0.272*					

An asterisk next to a parameter estimate indicates that it is constrained

The four-factor CFA Model and the MIRT both identified one item that showed significant misfit: F_Anx02, where students were asked to rate the activity of ñInterpreting statisticsò based on how much anxiety it gave them, which should be removed from the model (Table 19) ; however, in comparison when item thresholds (steps) are taken into account in Table 20, all items, including F_Anx02, appeared to have adequate fit based on infit (weighted) MNSQs (.75 and 1.33). Steps are used to account for partial credit, or in this case, partial agreement. Therefore, according to Table 20, when the interaction effect of the item levels of agreeability is controlled for, the SAM items no longer demonstrated a significant amount of misfit according to Wilsonôs (2005) standards (weighted MNSQs are greater than .75 and less than 1.33).

Table 20

Multidimensional Item Response Theory Response Model Parameter Estimates Based on Item Thresholds

TABLES OF RESPONSE MODEL PARAMETER ESTIMATES

TERM 2: item*step

VARIABLES				UNWGHTED FIT		WGHTED FIT	
item	step	ESTIMATE	ERROR	MNSQ	T	MNSQ	T
1 F_Fear01	1	-2.928	0.122	1.22	3.1	1.09	1.4
1 F_Fear01	2	0.438	0.156	1.73	8.9	1.02	0.2
2 F_Anx09	1	-2.185	0.120	0.75	-4.0	0.80	-3.5
2 F_Anx09	2	-0.002	0.118	0.80	-3.1	0.89	-2.4
3 F_Anx02	1	-2.207	0.118	1.09	1.3	1.06	1.0
3 F_Anx02	2	0.289	0.124	1.24	3.3	1.03	0.5
4 F_Anx04	1	-2.219	0.117	1.01	0.1	1.05	0.8
4 F_Anx04	2	0.074	0.131	1.02	0.3	0.96	-0.6
5 F_Anx05	1	-1.532	0.111	0.91	-1.4	0.95	-1.1
5 F_Anx05	2	0.034	0.138	0.99	-0.1	0.95	-0.8
6 F_Anx06	1	-1.537	0.111	1.02	0.3	1.04	0.8
6 F_Anx06	2	-0.103	0.132	1.09	1.3	1.03	0.5
7 F_Fear04	1	-2.563	0.115	1.03	0.5	1.04	0.7
7 F_Fear04	2	0.959	0.185	1.27	3.7	1.07	0.6
8 F_Class02	1	-3.235	0.139	0.77	-3.6	0.85	-1.9
8 F_Class02	2	0.555	0.126	1.21	2.9	1.00	-0.0
9 F_Class04	1	-2.223	0.125	0.86	-2.2	0.92	-1.2
9 F_Class04	2	-0.412	0.108	1.34	4.6	1.07	1.9
10 F_Class07	1	-3.078	0.136	0.87	-2.0	0.85	-2.0
10 F_Class07	2	0.125	0.126	0.87	-2.0	1.00	-0.1
11 F_Class08	1	-3.641	0.154	0.64	-6.1	0.84	-1.8
11 F_Class08	2	0.129	0.124	0.83	-2.7	0.99	-0.1
12 F_Class09	1	-3.179	0.137	1.24	3.3	1.10	1.2
12 F_Class09	2	0.411	0.128	2.28	14.0	1.08	1.4
13 F_Math01	1	-5.421	0.174	1.68	8.4	1.38	3.4
13 F_Math01	2	0.111	0.162	0.82	-2.9	1.20	2.2
14 F_Math02	1	-4.847	0.158	0.54	-8.3	0.86	-1.5
14 F_Math02	2	0.788	0.140	1.03	0.5	1.06	0.9
15 F_Math03	1	-4.762	0.162	0.85	-2.3	0.85	-1.6
15 F_Math03	2	0.409	0.141	0.84	-2.4	1.02	0.3
16 F_Math05	1	-5.278	0.174	0.89	-1.6	0.75	-2.6
16 F_Math05	2	0.334	0.154	0.91	-1.3	0.96	-0.5
17 F_Math08	1	-4.926	0.168	0.92	-1.1	0.79	-2.2
17 F_Math08	2	0.248	0.149	0.50	-9.1	0.84	-2.2
18 F_Math09	1	-4.685	0.165	0.70	-4.9	0.76	-2.6
18 F_Math09	2	0.034	0.142	0.87	-2.0	0.92	-1.0
19 F_Perf01	1	-5.203	0.161	1.42	5.5	1.02	0.3
19 F_Perf01	2	0.850	0.181	1.29	4.0	1.13	1.2
20 F_Perf02	1	-4.092	0.131	1.37	4.9	1.12	1.6
20 F_Perf02	2	0.730	0.219	4.39	28.1	1.07	0.5
21 F_Perf05	1	-4.277	0.141	0.71	-4.8	0.84	-2.0
21 F_Perf05	2	0.341	0.168	0.62	-6.6	0.92	-0.8
22 F_Perf06	1	-4.509	0.140	0.84	-2.5	1.02	0.2
22 F_Perf06	2	0.004	0.171	0.58	-7.2	1.01	0.1
23 F_Perf07	1	-4.377	0.142	0.75	-3.9	0.87	-1.6
23 F_Perf07	2	0.743	0.178	0.68	-5.3	0.93	-0.6

An asterisk next to a parameter estimate indicates that it is constrained

Furthermore, Wilson (2005) points out that although

> A knee-jerk reaction to finding evidence of misfit is to delete the items from the item set. é this may not be the best strategy for dealing with poor fit. First, the result could be due to a random fluctuationð after all, in a set of 20 parameter tests at Ŭ= .05, even if none of the parameters misfit, one would expect that one would come up as statistically significant just due to chance. é Second, the item that is showing poor fit may be a crucial one either because of its rarity in the item sample with respect to content or with respect to location.ò (p. 132)

In this case, the misfit of F_Anx02 could be random, since these results are based on one administration of the SAM. Further testing of the SAM would be required to determine if this estimate of misfit is reliable or not. More importantly though, F_Anx02 (Item 4) is in fact a rare item of the SAM in that its first threshold taps into a lower level of statistics anxiety that only one other item (from the anxiety factor) taps into (Figure 15).

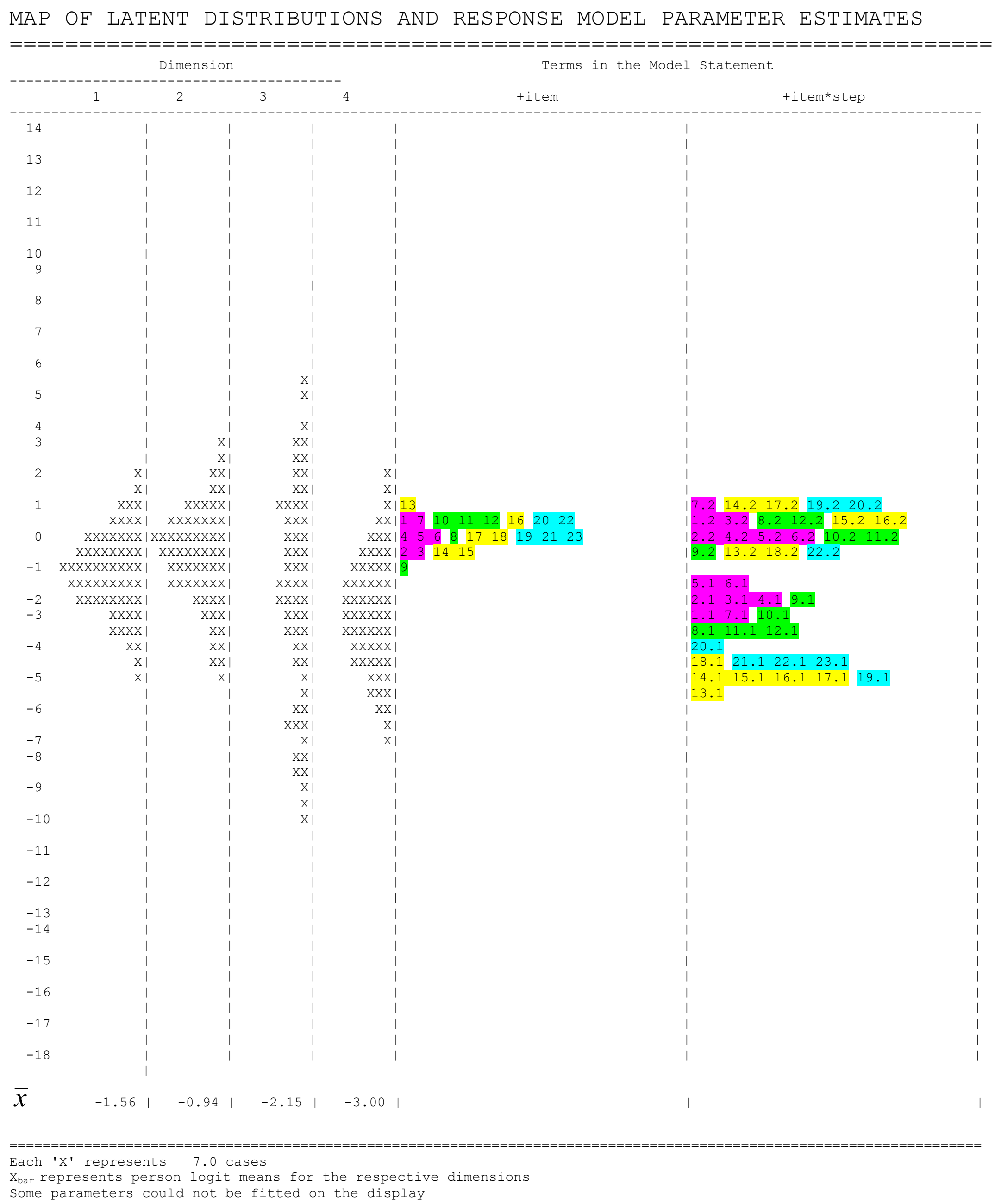

Figure 15. Final Administration Data Four-factor Item Difficulty Plot.

According to Figure 15, items were not widely dispersed; however, the persons sampled for this study appear to be less anxious on average on the anxiety factor, math factor, and performance factor than on the attitude towards class factor. Person anxiety in terms of class centered around a mean value of -0.64 logits, indicating that the anxiety of statistics class factor is doing a better job of representing all levels of anxiety as it pertains to the class than the other factors. Overall MIRT results suggest that the SAM be administered to students with higher levels of statistics anxiety or that items measuring lower levels of anxiety be developed.

Reliability

Cronbachȏs alpha was .93 for the 23-item scale with subscale reliabilities ranging from .82 to .95 (Table 21). According to Gable and Wolf (1993), good cognitive measures have alpha reliability ranging from the high .80s to the low .90s, but go on to say that good affective instruments frequently report reliabilities ranging in the 70s. Therefore, based on Cronbachȏs alpha it was found that the overall scale as well as the individual factor-based subscales all demonstrated adequate internal consistency levels.

Table 21

Internal Consistency Reliability of Four-factor Model Subscales (Cronbach's Alpha)

		Cronbach's α	N of Items
Scale	**SAM**	0.93	23
Subscales	**Anxiety**	0.86	7
	Class	0.82	5
	Math	0.95	6
	Performance	0.85	5

Invariance across groups could not be assessed using CFA, due to the limited size of the sample. With equal splits in any of the demographic groups for which data was collected, 552 cases would be needed to test invariance on a measure with 23 items. It is recommended in the future that the SAM be administered to a larger sample, consisting of at least 700 students, so that invariance across demographic groups may be assessed using CFA; however, the researcher was able to explore differential item functioning (DIF) using MIRT. According to Wu, Adams, and Wilson (1998), ñWithin the context of Rasch modeling, an item is deemed to exhibit differential item functioning (DIF) if the response probabilities for that item cannot be fully explained by the ability of the student and a fixed set of difficulty parameters for that itemò (p. 75). ConQuest was used to explore the existence of DIF with respect to gender, race (Caucasian vs. racial minorities), age (18-24 vs. 25-60), course level (undergraduate vs. graduate), and business versus nonbusiness statistics courses. This follows a traditional DIF approach since all of the variables listed above are scored dichotomously and thus focus on the comparison of two groups (Wu et al., 1998).

Three terms were used in the gender DIF MIRT model: item, gender, and item*gender. This model describes the probability of demonstrating greater anxiety on the items of SAM using an item main effect, a gender main effect, and an interaction effect of gender and item. The item main effect provides item difficulty estimates. The gender main effect provides mean logit positions for males and females. The interaction effect estimates the difference in item agreeability for SAM items based on gender and reflects potential item bias. Both the gender main effect and the interaction effect are of interest in exploring DIF across genders. Wu et al. explained that when a parameter estimate is more than double the standard error, it indicates a statistically significant difference exists between the two groups being tested. Therefore, according to Table 22, the main effect of gender was significant (estimate/std error = 0.032/0.010 = 3.2 > 2).

Table 22

DIF MIRT Model: Main Effect of Gender

TERM 2: gender

VARIABLES			UNWGHTED FIT		WGHTED FIT	
gender	ESTIMATE	ERROR	MNSQ	T	MNSQ	T
1 male	-0.032	0.010	1.04	0.5	0.99	-0.1
2 female	0.032*					

An asterisk next to a parameter estimate indicates that it is constrained

The difference between the male and female estimates equals .064 logits (0.032 ï (-0.032)), demonstrating that the female students scored .064 higher on the SAM than the male students. Although the main effect of gender indicates DIF may exist in the SAM, the magnitude of the DIF for gender appears minimal. If all items on the SAM expressed a difference estimate of .064 logits, the distribution of male ability would be shifted by 6.4% of a student standard deviation (Wu et al., 1998). According to Table 23, Item 7 (F_Fear04: ñI've avoided taking this class as long as possible.ò), showed the largest difference with .124 logits (0.062 ï (-0.062)) = 0.124), which is still only 12.4% of a student standard deviation.

Table 23

DIF MIRT Model: Interaction Effect of Gender by Item

TERM 3: item*gender

VARIABLES				UNWGHTED FIT		WGHTED FIT	
item	gender	ESTIMATE	ERROR	MNSQ	T	MNSQ	T
1 F_Fear01	1 male	-0.049	0.037	0.33	-9.7	0.30	-14.1
2 F_Anx09	1 male	0.012	0.036	0.39	-8.4	0.38	-12.6
3 F_Anx02	1 male	0.003	0.036	0.63	-4.4	0.56	-8.0
4 F_Anx04	1 male	0.034	0.037	0.59	-5.0	0.52	-8.7
5 F_Anx05	1 male	0.044	0.037	0.55	-5.6	0.51	-8.6
6 F_Anx06	1 male	0.018	0.037	0.57	-5.3	0.56	-7.6
7 F_Fear04	1 male	-0.062*					
8 F_Class02	1 male	-0.001	0.033	0.34	-9.4	0.33	-15.2
9 F_Class04	1 male	0.043	0.033	0.43	-7.7	0.41	-12.9
10 F_Class07	1 male	-0.019	0.033	0.31	-10.1	0.30	-16.1
11 F_Class08	1 male	-0.032	0.033	0.28	-10.7	0.28	-17.0
12 F_Class09	1 male	0.009*					
13 F_Math01	1 male	-0.031	0.037	0.28	-10.9	0.24	-16.7
14 F_Math02	1 male	-0.013	0.036	0.21	-12.7	0.16	-20.4
15 F_Math03	1 male	-0.054	0.036	0.20	-13.0	0.17	-20.1
16 F_Math05	1 male	0.009	0.036	0.17	-13.8	0.17	-19.8
17 F_Math08	1 male	0.033	0.036	0.12	-15.6	0.14	-21.6
18 F_Math09	1 male	0.056*					
19 F_Perf01	1 male	0.013	0.035	0.22	-12.2	0.19	-19.6
20 F_Perf02	1 male	-0.027	0.037	0.27	-11.0	0.25	-16.2
21 F_Perf05	1 male	-0.013	0.036	0.10	-16.9	0.09	-25.2
22 F_Perf06	1 male	0.012	0.036	0.23	-12.1	0.20	-18.8
23 F_Perf07	1 male	0.016*					
1 F_Fear01	2 female	0.049*					
2 F_Anx09	2 female	-0.012*					
3 F_Anx02	2 female	-0.003*					
4 F_Anx04	2 female	-0.034*					
5 F_Anx05	2 female	-0.044*					
6 F_Anx06	2 female	-0.018*					
7 F_Fear04	2 female	0.062*					
8 F_Class02	2 female	0.001*					
9 F_Class04	2 female	-0.043*					
10 F_Class07	2 female	0.019*					
11 F_Class08	2 female	0.032*					
12 F_Class09	2 female	-0.009*					
13 F_Math01	2 female	0.031*					
14 F_Math02	2 female	0.013*					
15 F_Math03	2 female	0.054*					
16 F_Math05	2 female	-0.009*					
17 F_Math08	2 female	-0.033*					
18 F_Math09	2 female	-0.056*					
19 F_Perf01	2 female	-0.013*					
20 F_Perf02	2 female	0.027*					
21 F_Perf05	2 female	0.013*					
22 F_Perf06	2 female	-0.012*					
23 F_Perf07	2 female	-0.016*					

An asterisk next to a parameter estimate indicates that it is constrained

Although the existence of DIF across genders is indicated by the significance of the main effect of gender, the magnitude seems minimal and therefore, overall the SAM appeared fairly invariant in terms of gender.

Three terms were used in the race DIF MIRT model: item, race, and item*race. This model describes the probability of demonstrating greater anxiety on the items of SAM using an item main effect, a race main effect, and an interaction effect of race and item. The item main effect provides item difficulty estimates. The race main effect provides mean abilities for racial minorities and Caucasians. The interaction effect estimates the difference in item agreeability for SAM items based on being Caucasian or not. Both the race main effect and the interaction effect are of interest in exploring DIF across racial minorities and Caucasians. According to Table 24, the main effect of race was not statistically significant (0.007/0.010 = .70 < 2) (Wu et al., 1998).

Table 24

DIF MIRT Model: Main Effect of Race

TERM 2: race

VARIABLES			UNWGHTED FIT		WGHTED FIT	
race	ESTIMATE	ERROR	MNSQ	T	MNSQ	T
1 noncaucasian	0.007	0.010	1.15	1.5	1.11	1.6
2 caucasian	-0.007*					

An asterisk next to a parameter estimate indicates that it is constrained

The difference between the noncaucasian and Caucasian estimates was .014 logits, demonstrating that the noncaucasian students scored .014 higher on the SAM than the Caucasian students. If all items on the SAM expressed a difference estimate of .014 logits, the distribution of noncaucasian ability would be shifted by 1.4% of a student standard deviation (Wu et al., 1998). Although the main effect of race does not specifically indicate DIF may exist in the SAM and the magnitude of the DIF for race appears minimal, there are a few items that that significantly vary (estimate/standard error > 2) based on being Caucasian or not: Items 3 (F_Anx02: ñInterpreting statisticsò), 7 (F_Fear04: ñI've avoided taking this class as long as possibleò), 9 (F_Class04: ñI am only taking this class because it is requiredò), and 11 (F_Class08: ñTaking this class will have little impact on my lifeò) (Table 25).

Table 25

DIF MIRT Model: Interaction Effect of Race by Item

TERM 3: item*race

item	race	ESTIMATE	ERROR	UNWGHTED FIT MNSQ	UNWGHTED FIT T	WGHTED FIT MNSQ	WGHTED FIT T
1 F_Fear01	1 noncaucasian	0.059	0.038	0.32	-9.9	0.29	-14.5
2 F_Anx09	1 noncaucasian	0.039	0.036	0.38	-8.6	0.36	-13.1
3 F_Anx02	1 noncaucasian	-0.127	0.036	0.58	-5.1	0.52	-8.8
4 F_Anx04	1 noncaucasian	-0.068	0.037	0.55	-5.6	0.50	-9.2
5 F_Anx05	1 noncaucasian	-0.033	0.037	0.52	-6.1	0.49	-9.1
6 F_Anx06	1 noncaucasian	0.044	0.037	0.56	-5.5	0.54	-8.1
7 F_Fear04	1 noncaucasian	0.086*					
8 F_Class02	1 noncaucasian	0.069	0.033	0.33	-9.5	0.33	-15.3
9 F_Class04	1 noncaucasian	0.104	0.033	0.43	-7.7	0.41	-12.9
10 F_Class07	1 noncaucasian	0.008	0.033	0.31	-10.0	0.30	-16.3
11 F_Class08	1 noncaucasian	-0.222	0.034	0.30	-10.3	0.29	-16.5
12 F_Class09	1 noncaucasian	0.041*					
13 F_Math01	1 noncaucasian	-0.048	0.037	0.26	-11.3	0.24	-17.1
14 F_Math02	1 noncaucasian	-0.011	0.036	0.19	-13.3	0.16	-20.7
15 F_Math03	1 noncaucasian	-0.008	0.036	0.19	-13.3	0.17	-20.2
16 F_Math05	1 noncaucasian	-0.004	0.036	0.16	-14.1	0.17	-20.2
17 F_Math08	1 noncaucasian	0.032	0.036	0.13	-15.4	0.13	-22.0
18 F_Math09	1 noncaucasian	0.040*					
19 F_Perf01	1 noncaucasian	-0.012	0.035	0.21	-12.6	0.19	-19.8
20 F_Perf02	1 noncaucasian	0.008	0.037	0.28	-10.7	0.25	-16.3
21 F_Perf05	1 noncaucasian	-0.010	0.036	0.09	-17.2	0.09	-25.3
22 F_Perf06	1 noncaucasian	0.014	0.036	0.22	-12.3	0.20	-19.0
23 F_Perf07	1 noncaucasian	-0.000*					
1 F_Fear01	2 caucasian	-0.059*					
2 F_Anx09	2 caucasian	-0.039*					
3 F_Anx02	2 caucasian	0.127*					
4 F_Anx04	2 caucasian	0.068*					
5 F_Anx05	2 caucasian	0.033*					
6 F_Anx06	2 caucasian	-0.044*					
7 F_Fear04	2 caucasian	-0.086*					
8 F_Class02	2 caucasian	-0.069*					
9 F_Class04	2 caucasian	-0.104*					
10 F_Class07	2 caucasian	-0.008*					
11 F_Class08	2 caucasian	0.222*					
12 F_Class09	2 caucasian	-0.041*					
13 F_Math01	2 caucasian	0.048*					
14 F_Math02	2 caucasian	0.011*					
15 F_Math03	2 caucasian	0.008*					
16 F_Math05	2 caucasian	0.004*					
17 F_Math08	2 caucasian	-0.032*					
18 F_Math09	2 caucasian	-0.040*					
19 F_Perf01	2 caucasian	0.012*					
20 F_Perf02	2 caucasian	-0.008*					
21 F_Perf05	2 caucasian	0.010*					
22 F_Perf06	2 caucasian	-0.014*					
23 F_Perf07	2 caucasian	0.000*					

An asterisk next to a parameter estimate indicates that it is constrained

Caucasians indicated a higher level of anxiety when interpreting statistics and were more likely to agree that they were only taking their statistics class because it was required than racial minorities were. Racial minority students were more likely to agree that they had avoided taking their statistics class as long as possible and that taking statistics would have little impact on their life than Caucasian students were. Not only did the probability of agreeing for the four items above vary significantly as a factor of being Caucasian or not, but the value of their difference logits is concerning. The logit position of racial minority students is shifted by 25.4% of a student standard deviation for Item 3, 17.2% for Item 7, 20.8% for Item 9, and 44% for Item 11. Although overall SAM may indicate invariance across Caucasians and racial minorities, there are items that should either be eliminated or addressed differently for Caucasians verses racial minorities when interpreting the results of the SAM.

Three terms were used in the age DIF MIRT model: item, age, and item*age. This model describes the probability of demonstrating greater anxiety on the items of SAM using an item main effect, an age main effect, and an interaction effect of age and item. The item main effect provides item difficulty estimates. The age main effect provides mean abilities for ages less than 25 and ages greater than 24. The interaction effect estimates the difference in item agreeability for SAM items based on age. Both the age main effect and the interaction effect are of interest in exploring DIF across students under the age of

25 and students 25 and over. According to Table 26, it appears the main effect of age was significant (0.068/0.010 = 6.8 > 2).

Table 26

DIF MIRT Model: Main Effect of Age

TERM 2: age

VARIABLES			UNWGHTED FIT		WGHTED FIT	
age	ESTIMATE	ERROR	MNSQ	T	MNSQ	T
1 <25	0.068	0.010	1.10	1.0	1.06	0.8
2 >24	-0.068*					

An asterisk next to a parameter estimate indicates that it is constrained

The difference between estimates for students under the age of 25 versus students 25 and over equaled .136 logits, demonstrating that the students under the age of 25 scored .136 higher on the SAM than students ages 25 and over. Although the main effect of age indicated DIF may exist in the SAM, the magnitude of the DIF for age was small. If all items on the SAM expressed a difference estimate of .136 logits, the distribution of ability for students under the age of 25 would be shifted by 13.6% of a student standard deviation (Wu et al., 1998). According to Table 27, the probability of agreement for Item 1 (F_Fear01: ñI was hesitant to register for this classò) and 7 (F_Fear04: ñI've avoided taking this class as long as possibleò) varied significantly as a result of being younger than 25 or 25 and older (See Table 27).

Table 27

DIF MIRT Model: Interaction Effect of Age and Items

TERM 3: item*age

VARIABLES				UNWGHTED FIT		WGHTED FIT	
item	age	ESTIMATE	ERROR	MNSQ	T	MNSQ	T
1 F_Fear01	1 <25	-0.079	0.038	0.32	-9.8	0.28	-14.8
2 F_Anx09	1 <25	-0.021	0.036	0.39	-8.3	0.37	-13.0
3 F_Anx02	1 <25	0.037	0.036	0.64	-4.3	0.54	-8.4
4 F_Anx04	1 <25	0.058	0.037	0.59	-5.0	0.50	-9.1
5 F_Anx05	1 <25	0.033	0.037	0.53	-6.0	0.49	-9.1
6 F_Anx06	1 <25	0.066	0.037	0.56	-5.4	0.53	-8.3
7 F_Fear04	1 <25	-0.094*					
8 F_Class02	1 <25	0.025	0.033	0.32	-9.7	0.32	-15.8
9 F_Class04	1 <25	-0.044	0.033	0.41	-8.0	0.40	-13.2
10 F_Class07	1 <25	0.012	0.033	0.31	-10.1	0.29	-16.7
11 F_Class08	1 <25	0.017	0.033	0.28	-10.7	0.27	-17.5
12 F_Class09	1 <25	-0.010*					
13 F_Math01	1 <25	-0.015	0.037	0.25	-11.4	0.24	-16.5
14 F_Math02	1 <25	-0.021	0.036	0.18	-13.4	0.16	-20.3
15 F_Math03	1 <25	-0.054	0.036	0.19	-13.3	0.17	-20.1
16 F_Math05	1 <25	0.000	0.036	0.18	-13.6	0.17	-19.7
17 F_Math08	1 <25	0.044	0.036	0.13	-15.5	0.14	-21.6
18 F_Math09	1 <25	0.046*					
19 F_Perf01	1 <25	-0.054	0.035	0.22	-12.2	0.19	-19.5
20 F_Perf02	1 <25	0.101	0.038	0.30	-10.2	0.27	-15.8
21 F_Perf05	1 <25	-0.054	0.036	0.09	-17.1	0.09	-25.0
22 F_Perf06	1 <25	0.068	0.037	0.24	-11.8	0.21	-18.4
23 F_Perf07	1 <25	-0.061*					
1 F_Fear01	2 >24	0.079*					
2 F_Anx09	2 >24	0.021*					
3 F_Anx02	2 >24	-0.037*					
4 F_Anx04	2 >24	-0.058*					
5 F_Anx05	2 >24	-0.033*					
6 F_Anx06	2 >24	-0.066*					
7 F_Fear04	2 >24	0.094*					
8 F_Class02	2 >24	-0.025*					
9 F_Class04	2 >24	0.044*					
10 F_Class07	2 >24	-0.012*					
11 F_Class08	2 >24	-0.017*					
12 F_Class09	2 >24	0.010*					
13 F_Math01	2 >24	0.015*					
14 F_Math02	2 >24	0.021*					
15 F_Math03	2 >24	0.054*					
16 F_Math05	2 >24	-0.000*					
17 F_Math08	2 >24	-0.044*					
18 F_Math09	2 >24	-0.046*					
19 F_Perf01	2 >24	0.054*					
20 F_Perf02	2 >24	-0.101*					
21 F_Perf05	2 >24	0.054*					
22 F_Perf06	2 >24	-0.068*					
23 F_Perf07	2 >24	0.061*					

An asterisk next to a parameter estimate indicates that it is constrained

Students ages 25 and older were more likely to agree with the above statements than students under the age of 25 (Table 27). The ability of students under the age of 25 is shifted by 15.8% of a student standard deviation for Item 1 and 18.8% for Item 7. DIF appears to be present when comparing students under the age of 25 with those ages 25 and older, specifically in terms of Items 1 and 7; therefore, these items should either be considered for elimination in the next study or their results should be interpreted differently based on age.

Three terms were used in the course level DIF MIRT model: item, course level, and item*course level. This model describes the probability of demonstrating greater anxiety on the items of the SAM using an item main effect, a course level main effect, and an interaction effect of course level and item. The item main effect provides item difficulty estimates. The course level main effect provides mean abilities for students in undergraduate courses and graduate courses. The interaction effect estimates the difference in item agreeability for SAM items based on course level. Both the course level main effect and the interaction effect were of interest in exploring DIF across students in undergraduate verses graduate courses. According to Table 28, the main effect of course level was significant (0.073/0.010 = 7.3 > 2).

Table 28

DIF MIRT Model: Main Effect of Course Level

TERM 2: graduate

VARIABLES			UNWGHTED FIT		WGHTED FIT	
graduate	ESTIMATE	ERROR	MNSQ	T	MNSQ	T
1 undergraduate	0.073	0.010	1.13	1.4	1.07	1.0
2 graduate	-0.073*					

An asterisk next to a parameter estimate indicates that it is constrained

The difference between estimates for students in undergraduate courses versus graduate courses equaled .146 logits, demonstrating that the students in undergraduate courses scored .146 higher on the SAM than students in graduate courses. Although the main effect of course level indicates DIF may exist in the SAM, the magnitude of the DIF for course was not extreme. If all items on the SAM expressed a difference estimate of .146 logits, the distribution of ability for students in undergraduate courses would be shifted by 14.6% of a student standard deviation (Wu et al., 1998). According to Table 29, the probability of agreement or indicating anxiety for Item 1 (F_Fear01: ñI was hesitant to register for this classò), 4 (F_Anx04: ñDeveloping conclusions based on mathematical solutionsò) and 7 (F_Fear04: ñI've avoided taking this class as long as possibleò) varied significantly as a function of course level.

Table 29

DIF MIRT Model: Testing the Interaction Effect of Course Level and Items

TERM 3: item*graduate

VARIABLES				UNWGHTED FIT		WGHTED FIT	
item	graduate	ESTIMATE	ERROR	MNSQ	T	MNSQ	T
1 F_Fear01	1 undergraduate	-0.079	0.038	0.32	-9.8	0.29	-14.5
2 F_Anx09	1 undergraduate	-0.024	0.036	0.38	-8.6	0.37	-13.0
3 F_Anx02	1 undergraduate	0.041	0.036	0.65	-4.1	0.54	-8.5
4 F_Anx04	1 undergraduate	0.085	0.037	0.55	-5.6	0.49	-9.3
5 F_Anx05	1 undergraduate	0.068	0.038	0.52	-6.1	0.48	-9.3
6 F_Anx06	1 undergraduate	0.038	0.037	0.58	-5.2	0.54	-8.2
7 F_Fear04	1 undergraduate	-0.129*					
8 F_Class02	1 undergraduate	0.005	0.033	0.34	-9.5	0.32	-15.5
9 F_Class04	1 undergraduate	-0.052	0.033	0.41	-7.9	0.40	-13.2
10 F_Class07	1 undergraduate	0.018	0.033	0.31	-10.1	0.30	-16.5
11 F_Class08	1 undergraduate	0.012	0.033	0.29	-10.5	0.28	-17.3
12 F_Class09	1 undergraduate	0.017*					
13 F_Math01	1 undergraduate	-0.001	0.037	0.26	-11.3	0.24	-16.9
14 F_Math02	1 undergraduate	0.010	0.036	0.21	-12.7	0.16	-20.6
15 F_Math03	1 undergraduate	-0.055	0.036	0.18	-13.6	0.17	-20.3
16 F_Math05	1 undergraduate	-0.018	0.036	0.16	-14.1	0.17	-20.0
17 F_Math08	1 undergraduate	0.023	0.036	0.13	-15.5	0.14	-21.9
18 F_Math09	1 undergraduate	0.040*					
19 F_Perf01	1 undergraduate	0.021	0.035	0.22	-12.4	0.19	-19.7
20 F_Perf02	1 undergraduate	0.029	0.037	0.29	-10.5	0.25	-16.2
21 F_Perf05	1 undergraduate	-0.019	0.036	0.10	-16.9	0.09	-25.2
22 F_Perf06	1 undergraduate	-0.030	0.036	0.23	-12.1	0.20	-18.9
23 F_Perf07	1 undergraduate	-0.001*					
1 F_Fear01	2 graduate	0.079*					
2 F_Anx09	2 graduate	0.024*					
3 F_Anx02	2 graduate	-0.041*					
4 F_Anx04	2 graduate	-0.085*					
5 F_Anx05	2 graduate	-0.068*					
6 F_Anx06	2 graduate	-0.038*					
7 F_Fear04	2 graduate	0.129*					
8 F_Class02	2 graduate	-0.005*					
9 F_Class04	2 graduate	0.052*					
10 F_Class07	2 graduate	-0.018*					
11 F_Class08	2 graduate	-0.012*					
12 F_Class09	2 graduate	-0.017*					
13 F_Math01	2 graduate	0.001*					
14 F_Math02	2 graduate	-0.010*					
15 F_Math03	2 graduate	0.055*					
16 F_Math05	2 graduate	0.018*					
17 F_Math08	2 graduate	-0.023*					
18 F_Math09	2 graduate	-0.040*					
19 F_Perf01	2 graduate	-0.021*					
20 F_Perf02	2 graduate	-0.029*					
21 F_Perf05	2 graduate	0.019*					
22 F_Perf06	2 graduate	0.030*					
23 F_Perf07	2 graduate	0.001*					

An asterisk next to a parameter estimate indicates that it is constrained

Students in graduate courses were more likely to agree that they were hesitant to register for the class and had avoided taking their class as long as possible than students in undergraduate courses (Table 29). Students in undergraduate courses had a higher probability of associating higher levels of anxiety with the activity of developing conclusions based on mathematical solutions than students in graduate courses (Table 29). The ability of students in undergraduate courses was shifted by 15.8% of a student standard deviation for Item 1, 17.0% for Item 4, and 25.8% for Item 7. DIF appeared to be present when comparing students in undergraduate courses with those in graduate courses, specifically in terms of Items 1, 4, and 7; therefore, these items should either be considered for elimination in the next study or their results should be interpreted differently based on course level.

Three terms were used in the business course DIF MIRT model: item, business course, and item*business course. This model describes the probability of demonstrating greater anxiety on the items of the SAM using an item main effect, a business course main effect, and an interaction effect of business course and item. The item main effect provides item difficulty estimates. The business course main effect provides mean abilities for students in business courses and nonbusiness courses. The interaction effect estimates the difference in item agreeability for SAM items based on the course being a business class or not. Both the business course main effect and the interaction effect are of interest in exploring DIF across students in nonbusiness verses business courses. According to Table 30, it appeared the main effect of business course was significant (0.044/0.010 = 4.4 > 2).

Table 30

DIF MIRT Model: Main Effect of Business Course

TERM 2: business

VARIABLES			UNWGHTED FIT		WGHTED FIT	
business	ESTIMATE	ERROR	MNSQ	T	MNSQ	T
1 nonbusiness	-0.044	0.010	1.12	1.2	1.08	1.1
2 business	0.044*					

An asterisk next to a parameter estimate indicates that it is constrained

The difference between estimates for students in nonbusiness courses versus business courses equaled .088 logits, demonstrating that the students in undergraduate courses scored .088 higher on the SAM than students in graduate courses. Although the main effect of business indicates DIF may exist in the SAM, the magnitude of the DIF for business course was minimal. If all items on the SAM expressed a difference estimate of .088 logits, the distribution of ability for students in nonbusiness courses would be shifted by 8.8% of a student standard deviation (Wu et al., 1998). According to Table 31, the probability of agreement or indicating anxiety for Item 7 (F_Fear04: ñI've avoided taking this class as long as possibleò) varied as a function of whether the course was business or non-business.

Table 31

DIF MIRT Model: Interaction Effect of Business Course and Items

TERM 3: item*business

item	business	ESTIMATE	ERROR	UNWGHTED FIT MNSQ	UNWGHTED FIT T	WGHTED FIT MNSQ	WGHTED FIT T
1 F_Fear01	1 nonbusiness	0.044	0.038	0.33	-9.7	0.29	-14.4
2 F_Anx09	1 nonbusiness	-0.013	0.036	0.38	-8.6	0.37	-12.9
3 F_Anx02	1 nonbusiness	-0.010	0.036	0.65	-4.2	0.55	-8.2
4 F_Anx04	1 nonbusiness	-0.036	0.037	0.57	-5.2	0.51	-8.9
5 F_Anx05	1 nonbusiness	-0.035	0.037	0.53	-6.0	0.50	-9.0
6 F_Anx06	1 nonbusiness	-0.044	0.037	0.55	-5.6	0.55	-8.0
7 F_Fear04	1 nonbusiness	0.094*					
8 F_Class02	1 nonbusiness	-0.033	0.033	0.33	-9.7	0.32	-15.6
9 F_Class04	1 nonbusiness	-0.012	0.033	0.41	-7.9	0.41	-13.1
10 F_Class07	1 nonbusiness	0.020	0.033	0.30	-10.2	0.30	-16.4
11 F_Class08	1 nonbusiness	0.012	0.033	0.29	-10.6	0.28	-17.2
12 F_Class09	1 nonbusiness	0.013*					
13 F_Math01	1 nonbusiness	0.004	0.037	0.28	-10.8	0.24	-16.8
14 F_Math02	1 nonbusiness	-0.018	0.036	0.20	-13.0	0.17	-20.5
15 F_Math03	1 nonbusiness	0.049	0.036	0.19	-13.3	0.17	-20.2
16 F_Math05	1 nonbusiness	0.022	0.036	0.18	-13.5	0.17	-20.0
17 F_Math08	1 nonbusiness	-0.021	0.036	0.13	-15.2	0.14	-21.8
18 F_Math09	1 nonbusiness	-0.036*					
19 F_Perf01	1 nonbusiness	-0.025	0.035	0.21	-12.6	0.19	-19.7
20 F_Perf02	1 nonbusiness	-0.025	0.037	0.28	-10.8	0.26	-16.1
21 F_Perf05	1 nonbusiness	0.006	0.036	0.09	-17.1	0.09	-25.1
22 F_Perf06	1 nonbusiness	0.045	0.036	0.22	-12.4	0.20	-18.8
23 F_Perf07	1 nonbusiness	-0.002*					
1 F_Fear01	2 business	-0.044*					
2 F_Anx09	2 business	0.013*					
3 F_Anx02	2 business	0.010*					
4 F_Anx04	2 business	0.036*					
5 F_Anx05	2 business	0.035*					
6 F_Anx06	2 business	0.044*					
7 F_Fear04	2 business	-0.094*					
8 F_Class02	2 business	0.033*					
9 F_Class04	2 business	0.012*					
10 F_Class07	2 business	-0.020*					
11 F_Class08	2 business	-0.012*					
12 F_Class09	2 business	-0.013*					
13 F_Math01	2 business	-0.004*					
14 F_Math02	2 business	0.018*					
15 F_Math03	2 business	-0.049*					
16 F_Math05	2 business	-0.022*					
17 F_Math08	2 business	0.021*					
18 F_Math09	2 business	0.036*					
19 F_Perf01	2 business	0.025*					
20 F_Perf02	2 business	0.025*					
21 F_Perf05	2 business	-0.006*					
22 F_Perf06	2 business	-0.045*					
23 F_Perf07	2 business	0.002*					

An asterisk next to a parameter estimate indicates that it is constrained

Students in nonbusiness courses were more likely to agree that they had avoided taking their class as long as possible than students in business courses (Table 31). The ability of students in nonbusiness courses was shifted by 18.8% of a student standard deviation for Item 7. DIF appeared to be present when comparing students in business courses with those in nonbusiness courses, specifically for Item 7; therefore, this item should either be considered for elimination in the next study or its results should be interpreted differently based on the course being business related or not.

Although DIF was indicated through either the main effect or specific items in all of the above models, some were of more concern than others. According to Table 32, Course Level and Age demonstrated the greatest amount of DIF, whereas Business Course, Race, and Gender demonstrated lesser amounts.

Table 32

Comparison of DIF MIRT Models

Comparision	Logit Differences	Estimate	Error	Estimate / Error
Gender	0.014	0.007	0.01	0.7
Race	0.064	0.032	0.01	3.2
Age	0.136	0.068	0.01	6.8
Course Level	0.146	0.073	0.01	7.3
Business Course	0.088	0.044	0.01	4.4

According to Table 33, gender, age, course level, and business course were all significantly correlated, and thus race was the only variable used to test

DIF that was independent of the other variables tested. The above items can be tested using a Pearson correlation since they consist of only two categories of classification.

Table 33

DIF Test Variable Correlations

Correlations

		Gender	Race	Age	Course Level	Business Class
Gender	Pearson Correlation	1				
	Sig. (2-tailed)					
	N	433				
Race	Pearson Correlation	.020	1			
	Sig. (2-tailed)	.676				
	N	433	433			
Age	Pearson Correlation	.220**	-.020	1		
	Sig. (2-tailed)	.000	.683			
	N	432	432	432		
Course Level	Pearson Correlation	.381**	-.086	.718**	1	
	Sig. (2-tailed)	.000	.074	.000		
	N	433	433	432	433	
Business Class	Pearson Correlation	-.350**	-.028	-.579**	-.756**	1
	Sig. (2-tailed)	.000	.558	.000	.000	
	N	433	433	432	433	433

**. Correlation is significant at the 0.01 level (2-tailed).

None of the groups tested for invariance showed extreme levels of DIF; however, one item in particular did: Item 7 (F_Fear04: ñI've avoided taking this class as long as possibleò) consistently demonstrated DIF across all five comparisons. According to Table 34, Item 7 resulted in shifts in the student standard deviation ranging from 12.4% to 44.0%.

Table 34

Comparison of Item 7 DIF across Gender, Race, Age, Course Level, and Business Course

Comparison	Logit Differences	Estimate	% of Shift in student standard deviation
Gender	0.124	0.062	12.40%
Race	0.440	0.222	44.00%
Age	0.188	0.094	18.80%
Course Level	0.258	0.129	25.80%
Business Course	0.188	0.094	18.80%

Females, Caucasians, students over the age of 24, students in graduate courses, and students in nonbusiness courses were more likely to agree that they had put off enrolling in their statistics course as long as possible; therefore, it is recommended that when the SAM is administered to groups varying in gender, race, age, course level, or course type that Item 7 be interpreted with caution.

Phase 4: Validation

Convergent validity was assessed though two correlational studies. The anxiety subscale of the SAM was expected to correlate highly with the STARS (.75 - .90), the attitude subscales of the SAM were expected to correlate highly with the SAT (.75 - .90), and all other subscales of the SAM were expected to correlate moderately with the STARS and the SATS (< .40) (Table 3). According to Table 35, the expected correlation between the anxiety subscale of the SAM and the STARS was met (.75); however, the SATS did not correlate highly with any of the subscales of the SAM or with the STARS total score, but it was significantly correlated with the anxiety and performance subscales of the SAM, SAM total score, and STARS total score.

Table 35

Correlations among SAM, SATS, and STARS Subscales and Total Scores

Correlations

		Anxiety Score	Class Score	Math Score	Performance Score	SAM Score	STARS Score	SATS Score
Anxiety Score	Pearson Correlation	1						
	Sig. (2-tailed)							
	N	433						
Class Score	Pearson Correlation	.425**	1					
	Sig. (2-tailed)	.000						
	N	433	433					
Math Score	Pearson Correlation	.582**	.360**	1				
	Sig. (2-tailed)	.000	.000					
	N	433	433	433				
Performance Score	Pearson Correlation	.543**	.373**	.533**	1			
	Sig. (2-tailed)	.000	.000	.000				
	N	433	433	433	433			
SAM Score	Pearson Correlation	.851**	.651**	.840**	.742**	1		
	Sig. (2-tailed)	.000	.000	.000	.000			
	N	433	433	433	433	433		
STARS Score	Pearson Correlation	.747**	.451**	.547**	.494**	.737**	1	
	Sig. (2-tailed)	.000	.000	.000	.000	.000		
	N	78	78	78	78	78	78	
SATS Score	Pearson Correlation	.324**	-.140	.140	.333**	.211*	.305*	1
	Sig. (2-tailed)	.001	.145	.145	.000	.027	.011	
	N	110	110	110	110	110	68	110

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

Although the researcher expected a stronger relationship between the SAM and the SATS, specifically with regard to the attitude subscales of the SAM (Class and Math), the lack of correlation between the SAM and the SATS and the presence of a strong correlation between the SAM and the STARS indicates that the SAM is more a measure of statistics anxiety than attitudes towards statistics. Consequently though, the lack of a significant correlation between the two attitude SAM subscales and the SATS raises questions about what these two subscales are actually measuring. It may be that the SAM and the SATS are tapping into different types of attitudes. The SATS measures attitudes towards statistics through the following four-factors: affect, cognitive competence, value, and difficulty, where the SAM is argued to measure attitudes toward the class and attitudes toward math as they relate to statistics anxiety. In fact, the STARS appears to tap into more similar types of attitudes in that it measures statistics anxiety using the following six factors: perceived worth of statistics, interpretation anxiety, test and class anxiety, conceptual self-concept, fear of asking for help, and fear of statistics teachers. The STARS was not only significantly related to the total score of the SAM, but it was significantly correlated with all the subscales comprising the SAM. The STARS was highly correlated only with the Anxiety subscale (.75) as was expected, but was more than moderately correlated (> .40) with the Class, Math, and Performance subscale scores. Since both the SAM and the STARS do not share high correlation levels with the SATS, this may indicate that attitudes towards statistics are comprised of a number of factors that none of the above measures are able to fully measure. Furthermore, it may also indicate that only certain types of attitudes towards statistics are highly correlated with statistics anxiety.

CHAPTER 4

DISCUSSION

A number of instruments exist for measuring statistics anxiety and attitudes towards statistics, all of which claim to be more valid and reliable than other instruments. To date little research if any had focused on how the research thus far could be integrated to create a more comprehensive measure of statistics anxiety. The research thus far had mainly focused on what distinguished these measures from each other. Additional models have mainly been used as validation tools, not as tools for researching what may be missing from one measure to the next. This dissertation was intended to add further insight into the construct of statistics anxiety by integrating prior work in the creation of the SAM. The main goal in creating the SAM was not to create yet another measure of statistics anxiety that could be argued as being better than those before; the main goal was to create a measure that encompassed all the competing theories regarding statistics anxiety, thus creating a fully comprehensive measure of statistics anxiety, which as a result in theory should provide a more thorough measure of statistics anxiety than existed previously.

Multidimensional measures of statistics anxiety and attitudes towards statistics have traditionally been evaluated using classical test theory (CTT) assessment methods such as exploratory factor analysis (EFA) and confirmatory factor analysis (CFA), but no studies were located where assessment of multidimensional measures of statistics anxiety or attitudes towards statistics were assessed using multidimensional item response theory (MIRT). In this research, constructs were assessed using both CTT and MIRT. Results of both approaches were compared and combined into a hybrid method for assessing both factor and item fit. First, items were assessed for factor fit using pilot data based on EFA item loadings. Second, item fit was assessed using pilot data based on weighted mean square error standardized *t* values. Third, item contribution to reliability was assessed using pilot data based on Cronbachôs alpha. Each item was reviewed using all three assessment methods, which resulted in the creation of a hybrid score. The hybrid score was created using the ratings of all three methods and was ultimately used to determine which items would be assessed using the final administration data. Assessment of the final administration data was done using both CTT and MIRT. The CTT model was tested using CFA, and MIRT was assessed using a multidimensional Rasch model.

The results of this dissertation suggest that statistics anxiety can be represented by perceived anxiety of class related activities, attitude towards class, attitude towards math, and expected performance on class related activities. In addition, a combined psychometric assessment approach using both CTT and

MIRT is possible, allowing psychometricians to reap the benefits of both approaches since the opportunity for exploring all of the following estimates is possible: item error variances, item mean square errors, item loadings, person loadings, item difficulty, person ability, invariance, differential item functioning, and item thresholds.

In this chapter, the results of the phases of measure construction are summarized and then discussed (planning, construction, quantitative evaluation, and validation). Noteworthy findings, limitations, and suggestions for further study follow.

Summary of Results

In Chapter one the purpose and the audience for the SAM were identified. In the creation of the SAM, the goal was to determine what the statistics anxiety and attitudes towards statistics research had in common and construct the SAM so that it measured the multiple dimensions of statistics anxiety discussed in the literature (Aiken, 1976; Benson, 1987; Benson & Bandalas, 1989; Breckler & Wiggins, 1989; Burton & Russell, 1979; Cruise & Wilken, 1980; Eagly & Chakin, 1992; Fishbein & Ajzen, 1975; Hendel, 1980; Richardson & Woolfolk, 1980; Rounds & Hendel, 1980; Smith, 1981; Sarason, 1980; Sutarso, 1992; Tobias, 1987; Zeidner, 1991; Zeidner & Safir, 1989), providing greater insight into statistics anxiety for statistics students, professors, and persons in charge of curriculum development.

Two pilot studies were carried out prior to the proposal of this dissertation where potential items for measuring fear of statistics were tested. Due the indistinguishable nature of statistics anxiety versus fear of statistics, two graduate students currently taking statistics who requested tutoring were interviewed and asked to differentiate between fear of statistics and statistics anxiety. While the results of these interviews led the researcher to believe that the two concepts were used interchangeably in the vernacular, through further research it became clear that the distinction between fear and anxiety lies in situations involving life and death circumstances vs. those that do not. Fear is used to distinguish those circumstances where death is a possibility, where anxiety is used to describe situations where death is not a likely outcome of failure. Once this distinction was made, terminology was altered to assess statistics anxiety as opposed to fear of statistics. The Fear of Statistics Test (FST) was replaced by the Statistics Anxiety Measure (SAM), which was developed through this dissertation. A four point rating scale carried over from the FST to the SAM, ensuring that persons could not remain indifferent to statements relating to statistics anxiety by being forced to agree or disagree with statements.

Research revolving around statistics anxiety and attitudes towards statistics were summarized through thematic analysis of constructs proposed by Aiken (1976), Benson (1987), Benson and Bandalas (1989), Breckler and Wiggins (1989), Burton and Russell (1979), Cruise and Wilken (1980), Eagly and Chakin (1992), Fishbein and Ajzen (1975) Hendel (1980), Richardson and

Woolfolk (1980), Rounds and Hendel (1980), Smith (1981), Sarason (1980), Sutarso (1992), Tobias (1987), Zeidner (1991), and Zeidner and Safir (1989). Inductive analysis identified six domains of statistics anxiety which were included in the original version of the SAM: anxiety, fearful behavior, attitude, expectations, history and self-concept, and performance. An item pool of 120 items (20 items per factor) intended to be distinguishable by both domain and level agreeability were created and reviewed by five experts and pretested through two cognitive interviews. The 10 items with the highest expert ratings of quality per factor were kept and the rest were dropped resulting in a total of 60 items being included in the pilot version of the SAM. These items were ordered within each factor based on expert rated agreeability levels. The interviews were used for clarifying wording of items as well as identifying repetitive items.

Sixty items were piloted during the first two weeks of Winter Quarter, 2007, resulting in a total pilot sample size of 347 students. It was originally intended that pilot items would be analyzed via CFA and MIRT; however, due to the limited sample size, the researcher was forced to use EFA in place of CFA to conduct initial item pool reduction. MIRT was used to reorder items based on the level of agreeability demonstrated in the pilot study. A six factor solution was tested using EFA; however, a sixth factor was not identifiable based on Klines (1998) two step rule. A five factor solution met rules for identification and resulted in a reduction of 17 items, providing a final administration version of the SAM with 43 items. Items loaded differently than expected in the EFA, which resulted in the redefining of the domains represented in the SAM: anxiety, attitude towards class, fearful behavior, attitude towards math, and performance.

Forty-three items were administered in a final administration of the SAM during the first two weeks of Spring Quarter 2007 resulting in a final administration sample size of 433 students. It was originally intended that all items of the final administration version of the SAM would be analyzed using CFA and MIRT; however, due to the limited sample size, the number of items had to be reduced before CFA could be done. Items were identified for retention or elimination using a hybrid score which took into account item EFA loadings, MIRT weighted mean square error *t* values of items, and reliability of scale if items were removed. Item reduction analyses were done using pilot data and results were applied to the final administration data in order to further reduce the item pool so that CFA could be conducted. The hybrid score indicated 24 items should be kept based on factor membership and level of agreeability. These 24 items were analyzed using the five factor solution provided in the EFA of the pilot data; however, two factors were collapsed so that each factor was represented by a minimum of 5 items each. The anxiety and fearful behavior factors were deemed the most similar of the subscales and thus were combined, which resulted in two items measuring the same level of agreeability; the item with the lowest hybrid score was removed, resulting in a total of 23 items being included in the CFA of the four factor solution using the final administration data. Although the four factor solution demonstrated significantly more misfit than the five factor solution, the fit remained fair.

Reliability assessment of the reduced item pool using the final administration data demonstrated overall scale (Ŭ= 0.93) as well as subscale internal consistency (Ŭranged from 0.82 to 0.95). Invariance of items could not be tested via CFA due to the limited sample size. Differential item functioning (DIF) of gender, race (Caucasian vs. racial minorities), age (18-24 vs. 25-60), course level (undergraduate vs. graduate), and business vs. nonbusiness statistics courses was explored via MIRT. Gender, race, age, course level, and business vs. nonbusiness course classification all demonstrated DIF, but to varying degrees; Course level and age demonstrated greater amounts of DIF than business vs. nonbusiness courses, race, or gender. None of the five variables tested demonstrated extreme levels of DIF, but one item in particular consistently demonstrated DIF in all five comparisons; therefore, it was recommended that agreement with the statement ñIôve put off taking statistics as long as possibleò be interpreted with caution when administering the SAM to groups varying in gender, race, age, course level, or course type.

Convergent validity of the SAM was also assessed through administration of two additional measures in a limited number of classes. The anxiety scale of the SAM was expected to highly correlate (*r* Ó.75) with the Statistics Anxiety Rating Scale (STARS: Cruise & Wilkins, 1980) and the two attitude subscales of the SAM were expected to highly correlate with the Survey of Attitudes Towards

Statistics (SATS: Schau et al., 1995). The anxiety subscale did highly correlate with the STARS as expected; however, neither attitude subscale of the SAM highly correlated with the SATS leading the researcher to believe that only certain types of statistics related attitudes highly correlated with statistics anxiety.

Discussion of Results

The SAM was intended to be an all-encompassing measure of statistics anxiety and its influence on attitudes towards statistics counterparts. Measures thus far have focused on either statistics anxiety or attitudes towards statistics, but none to date have integrated the two to create a fully comprehensive structure of statistics anxiety. The structure laid out by Wiseøs (1985) Attitudes Towards Statistics measure and the anxiety structures laid out by Cruise and Wilkenøs (1980) STARS and Zeidnerøs (1991) Statistics Anxiety Inventory (SAI) are all captured in the four-factor structure of the SAM, which includes items measuring attitudes towards the course (Wise), attitudes towards the field (Wise), worth of statistics (Cruise & Wilken), interpretation anxiety (Cruise & Wilken), class anxiety (Cruise & Wilken), conceptual self concept (Cruise & Wilken), statistics content anxiety (Zeidner) and statistics test anxiety (Zeidner).

The SAM was originally intended to measure six dimensions of statistics anxiety (anxiety, fearful behavior, attitudes, expectations, history and self-concept, and expected performance); however, due to lack of empirical evidence and a limited sample size, the number of factors were reduced and restructured in the process to represent the following constructs instead: anxiety, attitude towards class, attitudes towards math, and expected level of performance. Items measuring fearful behavior, expectations, and history and self-concept were either reassigned based on empirical results or deleted entirely. The possibility remains that the original construct may exist, but the item pool failed to capture it. On the other hand, there is evidence that fearful behavior, expectations, and history and self-concept are actually captured in the resulting four dimensions of the SAM considering items intended to measure each of these constructs were contained in the final version of the SAM. The final version of the SAM resulted in a broad measure of statistics anxiety with a simple structure containing distinguishable subscales. The resulting subscales are easy to interpret and are useful in determining the type of activities, attitudes, and types of performance assessment responsible for statistics anxiety among students, which in the end was the ultimate goal in developing the SAM.

The SAM allows students, teachers, and researchers to assess levels of statistics anxiety for individuals, classes, and other adult populations. Understanding the level of statistics anxiety existing within an individual, a class, or a population, provides greater understanding for how topics should be introduced as well as indicating the type of activities that may encourage or discourage students to further their understanding of statistics. Using the multiple subscales of the SAM, professors have the opportunity to explore the extent to which their students are anxious, as well as to assess which areas that are contributing most to their anxiety. For example, some students may be confident in their mathematical abilities, but lack confidence in their ability to perform; other students may express little anxiety toward statistical activities, but express a negative attitude towards the class. Understanding studentsô preconceptions allows one to anticipate areas where they are most likely to lose students along the way. If students express negative attitudes towards math, attention might be focused towards reviewing certain mathematical principles as they apply to statistics through activities geared toward improving attitudes toward math and alleviating math anxiety: desensitizing, group therapy, math immersion, creative thought, and humor. According to Tobias (1978), programs employing traditional counseling techniques have been successful in reducing student levels of math anxiety and improving attitudes towards math; however, this technique requires a great deal of time and effort on the part of the participants as well as resources from the respective department. Another approach recommended by Schacht and Stewart (1990) for alleviating math and statistics anxiety and improving attitudes towards math, statistics, and the course involves the use of humor in the teaching and examination of statistics. Specifically they recommend basing lessons on statistical studies that encourage laughter. For example, Schacht and Stewart used a comic *Bloom County* depicting a nerdy looking man placing a personals ad asking his cat to please return home. They then asked students to estimate probabilities of runaway pets based on data they made up in connection with the comic. Although they were originally concerned about students not taking activities seriously due to the absurdity of their applications, they found that students were more engaged since they were able to find humor not just in the comics but also in the activities they were asked to carry out.

When comparing individual and classroom level data, professors can gain insight using the anxiety and the performance subscale of the SAM. The anxiety subscale can be used as a tool for anticipating the activities in the course that are likely to increase anxiety levels in individuals/classes. For example, on the SAM students may indicate feelings of anxiety when calculating probabilities; the instructor can use this information to allocate more time and thus attention to the calculation of probabilities, which hopefully in turn will reduce students' anxiety. The performance factor provides insight as to how to best assess student achievement across individuals, different departments, and course levels. For example, some students/classes may specify that they expect to perform better on class projects and quizzes than exams. In this case a professor might want to consider administering multiple quizzes and requiring a final project as opposed to exams. This method can be applied as an either/or option at the student level or as an overall change in assessment practice at the classroom level or even the departmental level. The SAM is not expected to be used as a tool for individualizing instruction, although it could be; the SAM is intended to be used as a tool for instructors to reevaluate their teaching and assessment methods in given populations of students (i.e. graduate versus undergraduate students, business versus psychology students, etc.).

The SAM was created based on the experiences this researcher had as a statistics student, instructor, tutor, and academic counselor for students with learning disabilities. The items within this measure give students the opportunity to pinpoint what makes them anxious and/or they dislike about statistics, while providing professors with insight as to where such statistics anxiety or distain come from. The SAM provides researchers with the tools for understanding relationships among statistics activities, anxiety, attitudes, and performance expectations, which in turn can provide further insight to statistics professors, departments, and universities as a whole to where efforts should be focused when working to alleviate statistics anxiety and thus improve attitudes toward the course as a whole.

Methodological Conclusions

According to Kline (1998) structural models of cognitive measures should be theory driven; however, with limited sample sizes, CFA oftentimes ends up being driven by numbers. A certain sample size is required in order to assess more complex models, so when the sample size falls short of that, the final structure of the cognitive measures may rely more on estimates than theory. Based on the results of this dissertation, MIRT is better equipped for dealing with smaller sample sizes than CFA. Unlike CFA, estimation in MIRT is not limited by sample size.

Statistics anxiety may in fact be best represented by a six factor solution; however, in the course of this dissertation the originally proposed structure did not hold given the items created by the researcher. Invariance could not be tested via CFA and thus compared with DIF results obtained via MIRT due to the limited sample size, thus larger sample sizes are recommended in future studies since such comparisons would be interesting considering the lack of research currently available where MIRT models are tested for DIF using ConQuest (Wu et al., 1998).

Research is recommended at institutions beyond the University of Denver (DU), since research involving the SAM thus far can only be generalized within statistics course populations at DU. Without further research it is unknown whether findings documented in this dissertation will hold for other populations.

Limitations

The theory of MIRT thus far is limited. No books exist specifically addressing this topic other than the manual created to accompany the ConQuest software (Wu. Et al, 1998). MIRT currently exists at best as a chapter within books on IRT and at worst as nothing more than a paragraph. A number of studies exist where MIRT was used, but they all vary in their method. The majority of published research relies on the interpretation of traditional IRT models and comparisons with CTT models using EFA and CFA. Without a specific protocol for assessing MIRT models, the burden of reinventing and testing MIRT models relies on researchers.

The items used to construct the SAM were written solely by the author of this dissertation, and were developed under time constraints, so there was little to no time for discussion with fellow researchers of their actual content other than the expert review where items were only evaluated empirically. Items did not load as theoretically expected, and thus the structure evolved through empirical findings of EFA and constraints of CFA. Given more time, more item pool creation phases would have been implemented as well as focus groups for the discussion of items using a mixture of professors, students, tutors, and psychometricians.

The sample used for the SAM assessment although broad within the DU setting, was limited as a whole. The largest sample consisted of 433 students which placed considerable constraints on the number of items that could be used in testing the structure of the SAM via CFA. Furthermore, due to the limited sample size, structures found in this study may not hold in future administrations given larger sample sizes.

Another concern is that the sample itself was limited solely to DU students, which provides its own biases. DU is a small private liberal arts university consisting of approximately 10,823 students, 5,122 undergraduate and 5,701 graduate students. The majority of undergraduate students admitted to DU in 2006 were female (52%) and/or Caucasian (85.7%), with median GPAs ranging from 3.38 to 3.96. Tuition at DU for the year 2006-2007 was $823 a credit hour, with mandatory housing and meal plan costs ranging from approximately $6,000 to $10,000 for an academic year. During the 2006-2007 academic year, less than half of undergraduate students were awarded need based financial aid (43%). The average undergraduate class size provided one instructor for every ten students. The majority of graduate students at DU, are enrolled in masters program (71%) and are full time students (55%). The average age of graduate students at the DU is 26. Results obtained from data collected at DU may not be generalizable to larger more diverse public colleges, especially non-liberal arts colleges.

Conclusion

Although limitations of the SAM have yet to be fully assessed, the SAM provides the most theoretically comprehensive and useful measure of statistics anxiety to date, combining all prior research on statistics anxiety and attitudes towards statistics while pinpointing areas of concern for students, teachers, tutors, academic counselors, and researchers. The SAM is a tool that students can use to determine where their anxiety of statistics comes from, teachers can use to determine how to structure their class, academic counselors can use to determine which style of instruction will best fit their student, and researchers can use when further exploring relationships with and within statistics anxiety.

The SAM provides students with a tool for identifying triggers of their statistics anxiety, and in turn professors with a diagnosis of their greatest hurdles for improving the statistical literacy of their students. The SAM bridges a communication gap between students and teachers; students are given the opportunity to identify anxiety provoking activities, explore as well as express attitudes towards the class and math, and determine performance expectations for a variety of assessment methods; teachers are in turn able to understand which activities students are most anxious about, studentsô attitudes towards the class and math, and studentsô performance expectations on a variety of assessments.

The SAM provides teachers with insight at the individual, classroom, departmental, and institutional level to the most anxiety provoking activities, influential attitudes towards the class or math, and areas with low performance expectations. This kind of insight is essential for increasing statistical literacy. An understanding of the most anxiety provoking activities in a statistics class, better prepares teachers for allocating time so it best suits the needs of their students. An understanding of student attitudes towards the class, reminds teachers of the importance of demonstrating the value of statistical literacy and the impact it has on oneøs life. An understanding of student attitudes towards math helps teachers identify their audience and determine the extent to which their efforts should focus on reviewing mathematical principles and improving studentsô math confidence. An understanding of student performance expectations, allows professors to determine how to best assess student ability for a given audience. Through use of the SAM, students and teachers are able to work together towards improving statistics courses and increasing statistical literacy.

Suggestions for Further Research

As stated earlier, it is necessary that the SAM be administered in other populations in order to determine how generalizable the construct is. Some recommendations include: public colleges, technical colleges, colleges catering to nontraditional students, community colleges, foreign universities, and public and private high schools including advanced placement statistics classes vs. non-advanced placement classes. Furthermore, it is necessary that higher order effects of statistics anxiety be assessed via hierarchical linear modeling across classroom, teacher, departmental, and institutional levels to determine if initial SAM scores and rates vary as a function of higher order effects.

It is recommended that the SAM be administered to a larger sample size to determine if the structure holds using CFA. A larger sample size is also recommended so invariance can be assessed using CFA and compared with DIF results obtained in MIRT analysis. It is recommended that future administrations of the SAM include samples of at least 700 cases since at a minimum 552 complete cases are required to test invariance between groups on a measure of 23 items.

In order to ensure the empirical findings of this study reflect the structure theorized in the construction of the SAM, it is recommended that further analysis be done using Figure 1 and that focus groups be conducted in order to determine potential factors and/or items that are missing in this comprehensive measure of statistics anxiety.

Lastly, efforts should be made to standardize and document the process in which MIRT and CFA can be best used in combination so that their findings fully complement the short comings of the other. Ideal software would assess structure, item fit, and person fit without incorrectly assuming that data are interval; Furthermore, ideal software would recommend items and factors for removal based on model, item, and person fit estimates while also considering issues of invariance such as differential item functioning in order to meet structural identification requirements and further ensure the generalizability of results across different demographic and educational groups.

REFERENCES

The American Heritage Dictionary (3rd ed.) (1994). New York: Houghton Mifflin Company.

Aiken, L. R. (1976). Update on Attitudes and other Affective Variables in Learning Mathematics. *Review of Educational Research, 46,* 293-311.

Auster, C. J. (2000). Probability Sampling and Inferential Statistics: An Interactive Exercise Using M&M's. *Teaching Sociology, 28*(October), 379-385.

Benson, J. (1987). Causal Components of Test Anxiety in Adults. Paper presented at the annual meeting of the Society for Test Anxiety Research, Bergen, Norway, June.

Benson, J., & Bandalos, D. (1989). Structural Model of Statistics Test Anxiety. In Schwarzer, R., Van Der Ploeg, H., and Spielberger, Ch.D. (Eds.), *Advances in Test Anxiety Research Vol. 6.* Lisse, Netherlands: Swets and Zeitlinger, Hillsdale, N.J.: Erlbaum.

Bond, T. G., & Fox, C. M. (2001). *Applying the Rasch Model: Fundamental Measurement in the Human Sciences.* Mahwah, New Jersey: Lawrence Erlbaum Associates.

Breckler, S. J., & Wiggins, E. C. (1989). Affect Versus Evaluation in the Structure of Attitudes. *Journal of Experimental Social Psychology, 25*, 253-271.

Bryan, J. H., Sonnefeld, L. J., & Grabowski, B. (1983). The Relationship Between Fear of Failure and Learning Disabilities. *Learning Disability Quarterly, 6*(2), 217-222.

Burton, G. M., & Russel, D. (1979). Getting Comfortable with Mathematics. *Elementary School Journal 79*, 129-135.

Byrne, B. M. (2001). *Structural Equation Modeling with AMOS.* Mahwah, New Jersey: Lawrence Erlbaum Associates.

Connolly, A. (1989). Taking the Fear Away from Learning. In *Children for the Learning and Teaching of Literature: Vol. 2.3. Observations of Contemporary Fiction Taught by Carol Hill School in a City in New York* (pp. 2-26). Albany: University at Albany State University of New York School of Education.

Cruise, R. J., Cash, R. W, & Bolton, D. L. (1985,August). Development and validation of an instrument to measure statistical anxiety. Paper presented at the annual meeting of the Statistical Education Section Proceedings of the American Statistical Association.

Cruise, R. J., & Wilkins, E. M. (1980). STARS: Statistical Anxiety Rating Scale. Unpublished manuscript, Andrews University, Berrien Springs, MI.

DeVellis, R. F. (1991). *Scale Development: Theory and Applications: Vol. 26. Applied Social Research Methods Series.* Newbury Park: Sage Publications.

Eagly, A. H., & Chaiken, S. (1992). *The Psychology of Attitudes*. San Diego, CA: Harcourt Brace Jovanovich.

Earp, M. S. R. (2003). Fear of Statistics Measure. In a *University of Denver Course: Structural Equation Modeling*. Pilot study of five introductory statistics classes at the University of Denver.

Embretson, S. E., & Reise, S. P. (2000). *Item Response Theory for Psychology.* Mahwah, New Jersey: Lawrence Erlbaum Associates.

Fishbein, M., & Ajzen, I. (1975). *Belief, Attitude, Intention, and Behavior: An Introduction to Theory and Research.* Reading , MA: Addision-Wesley.

Gable, R. K., & Wolf, M. B. (1993). *Instrument Development in the Affective Domain* (Second ed.). Boston: Kluwer Academic Publishers.

Green, K. E. (1993, April). Affective, Evaluative, and Behavioral Components of Attitudes Toward Statistics. In *Statistics Students, Their Attitudes, Achievement, and Related Characteristics.* Symposium conducted at the meeting of the American Educational Research Association Annual Meeting, Atlanta.

Green, K. E. (1994, April). The Affective Component of Attitude in Statistics Instruction. In *Current Trends and Issues in Teaching Statistics.* Symposium conducted at the meeting of the American Educational Research Association, New Orleans.

Hendel, D. D. (1980). Experimental and Affective Correlates of Math Anxiety in Adult Women. *Psychology of Women Quarterly, 5*, 219-289.

Jordan, J., & Haines, B. (2003). Fostering Quantitative Literacy. *Peer Review, 5*(4), 16-19.

Jöreskog, K., & Sörbom, D. (1993). *LISREL 8: Structural Equation Modeling with the SIMPLIS Command Language.* Lincolnwood, Illinois: Scientific Software International.

Kline, R. B. (1998). *Principles and Practice of Structural Equation Modeling.* New York: The Guilford Press.

MacCallum, R., Widaman, K., Preacher, K., & Hong, S. (2001, October). Sample Size in Factor Analysis: The Role of Model Error. *Multivariate*

Behavioral Research, *36*(4), 611-637. Retrieved May 27, 2007, from Academic Search Premier database.

Mji, A., & Onwuegbuzie, A. J. (2004). Evidence of Score Reliability and Validity of the Statistical Anxiety Rating Scale Among Technikon Students in South Africa. *Measurement and Evaluation in Counseling and Development, 36,* 238-251.

Nemours Foundation. (2002, April). Fears and Phobias [What's the Difference Between Fear and Anxiety] D. Lyness (Ed.),. Article posted to KidsHealth: http://www.kidshealth.org/teen/your_mind/mental_health/phobias_p2.html

Patton, M. Q. (2002). *Qualitative Research & Evaluation Methods* (Third ed.). Thousand Oaks: Sage Publications. (Original work published 1990)

Pretorius, T. B. & Norman, A. M. (1992). Psychometric data on the Statistics Anxiety Scale for a sample of south African Students. *Educational and Psychological Measurement, 52*(4), 933-937.

Richardson, F. C., & Woolfolk, R. L., (1980). Mathematics Anxiety. In Sarason, I. G. (Ed.), *Test Anxiety: Theory, Research, and Applications,* Hillsdale, N.J.: Erlbaum.

Roberts, D. M., & Bilderback, E. W. (1980). Reliability and Validity of Statistics Attitude Survey. *Educational and Psychological Measurement, 40,* 235-238.

Roberts, D. M., & Reese, C. M. (1987). A Comparison of Two Scales Measuring Attitudes Toward Statistics. *Educational and Psychological Measurement, 47,* 759-764.

Roberts, D. M., & Saxe, J. E. (1982). Validity of a Statistics Attitude Survey: A Follow-Up Study. *Educational and Psychological Measurement, 42,* 907-912.

Rounds, J. B., & Hendel, D. D. (1980, March). Measurement and Dimensionality of Mathematics Anxiety. *Journal of Counseling Psychology, 27*(2), 138-149.

Sahai, H., & Khurshid, A. (2002). *Pocket Dictionary of Statistics.* Boston: McGraw Hill.

Sarason, I.G. (1980). *Test Anxiety: Theory, Research and Applications.* Hillsdale, N.J.: Erlbaum.

Schacht, S., & Steward, B. J. (1990). What's Funny About Statistics? A Technique for Reducing Student Anxiety. *Teaching Sociology, 18*(January), 52-56.

Schau, C., Stevens, J., Dauphinee, T. L., & Del Vacchio, A. (1995). The Development and Validation of the Survey of Attitudes Toward Statistics. *Educational and Psychological Measurement, 55*(5), 868-875.

Schield, M. (2004). Statistical Literacy and Liberal Education at Augsburg College. *Peer Review, 6*(4), 16-18.

Smith, J. B. (1981). *Math Anxiety and the Student of the '80s.* ERIC Documentation Reproduction Service No. ED, 200-296.

Sutarso, T. (1992, November). Some Variables in Relation to Studentsô Anxiety in Learning Statistics. Paper presented at the annual meeting of the Mid-South Educational Research Association, Knoxville.

Tobias, S. (1978). *Overcoming Math Anxiety*. New York: Norton.

Tobias, S. (1987). Math Anxiety. *Science, 237*, 1556.

Twain, M. (1924). *Mark Twain's Autobiography.* Harper & Brothers Publishers.

Value Quotes. (n.d.). Retrieved October 28, 2006, from http://www.valuequotes.net/

Waters, T. A., Martelli, T. Z., & Popovich, P. M. (1988). Attitudes Toward Statistics: An Evaluation of Multiple Measures. *Educational and Psychological Measurement, 48,* 513-516.

Webster's New World Dictionary (College ed.) (1957). Cleveland: The World Publishing Company.

Williams, K. (2001). Understanding, Communication Anxiety, and Gender Physics. *Journal of College Science Teaching, 30*(4), 232-237.

Wilson, M. (2005). *Constructing Measures: An Item Response Modeling Approach.* Mahwah, New Jersey: Lawrence Erlbaum Associates.

Wise, S. L. (1985). The Development and Validation of a Scale Measuring Attitudes Toward Statistics. *Educational and Psychological Measurement, 45,* 401-405.

Wu, M. L., Adams, R. J., & Wilson, M. R. (1998). *ACER ConQuest: Generalized Item Response Modeling Software.* Melbourne, Victoria: Australian Council for Educational Research.

Zeidner, M. (1991). Statistics and Mathematics Anxiety in Social Science Students: Some Interesting Parallels. *British Journal of Educational Psychology, 61,* 319-328.

Zeidner, M., & Safir, M. (1989). Sex, Ethnic, and Social Differences in Test Anxiety among Israeli Adolescents. *Journal of Genetic Psychology, 150*, 175-185.

APPENDIX A

Expert Review Request Letter

Expert Review Letter:

Dear [insert faculty/staff person's name],

Greetings! My name is Morgan Earp. I am a doctoral candidate in the College of Education's Quantitative Research Methods program. I am developing and validating a measure of statistics anxiety in partial fulfillment of the degree, doctorate of philosophy. As an undergraduate transfer student in 1999, I was required, as were all students, to take introduction to statistics. As a graduate student in the Graduate School of International Studies, I was required to take statistical methods. It was through these experiences that my interest in the topic of statistics anxiety began. Throughout my college experience I struggled with every class, that is until I discovered statistics. What intrigued me is that for the majority of my peers, the situation was opposite mine. Those who were fully confident of their academic abilities suddenly doubted their ability to succeed when it came time to take statistics. As I am sure you are aware the general public shares a distaste for this subject and thus many of our students hold negative views and/or doubt their ability to succeed before our class even starts [peers hold negative views and/or doubt their ability to understand our research findings].

I have spent the last four years researching the factors that comprise statistics anxiety. I am now moving into the final phase of my research at DU, and I am hoping for your participation in the expert review of my measure of statistics anxiety. The Statistics Anxiety Measure item pool has 6 domains with 20 items each. The review should take no longer than 30 minutes. You will be asked to complete an online survey rating of the quality of each item, appropriateness of domain, and the perceived level of agreeability for that item using an ordinal scale.

The results of this study will be available online at http://www.du.edu/~mearp/SAM. Upon graduation, the researcher will update the webpage with a direct link to the published dissertation. I intend that the results of this study will benefit you the instructor by providing a review of the factors that correlate with our students suffering statistics anxiety.

Please respond to this email, and let me know if you are willing to participate in the expert review of my measure. If you choose to participate, I will respond by sending you the link to the online survey.

Thank you for you time and support! I look forward to hearing from you.

Morgan S. Earp

(720) 217-5824

APPENDIX B

Cognitive Interview Consent Form for Statistics Anxiety Measure Study

Cognitive Interview Consent Form:

Dear [statistics studentŝ name],

Greetings! My name is Morgan Earp. I am a doctoral candidate at the University of Denver in the College of Education's Quantitative Research Methods program. I am developing and validating a measure of statistics anxiety in partial fulfillment of the degree, doctorate of philosophy. As an undergraduate transfer student in 1999, I was required, as were all students, to take introduction to statistics. As a graduate student in the Graduate School of International Studies, I was required to take statistical methods. It was through these experiences that my interest in the topic of statistics anxiety began. Throughout my college experience I struggled with every class, that is until I discovered statistics. What intrigued me is that for the majority of my peers, the situation was opposite mine. Those who were fully confident of their academic abilities suddenly doubted their ability to succeed when it came time to take statistics. As I am sure you are aware the general public shares a distaste for this subject and thus many students hold negative views and/or doubt their ability to succeed before their statistics class even begins.

I have spent the last four years researching the factors that comprise statistics anxiety. I am now moving into the final phase of my research at the University of Denver, and I am hoping to conduct a cognitive interview with you to assess the validity of my item pool. The Statistics Anxiety Measure item pool has 6 domains with 20 items each. The interview should take no longer than 1 hour. You will be asked to read each question and think aloud. The goal is for me the researcher to understand how you are interpreting the question at first glace. This will allow me to determine if the items are being interpreted they way they were intended to be. Your interview is confidential and your name will not be used in any reports.

This study is being conducted by myself, Morgan Earp, with supervision by Dr. Kathy Green, College of Education, 303-871-2490, kgreen@du.edu, whom you may contact if you wish further information or if you have any concerns about the project. If you have any concerns or complaints about how you were treated during the process, please contact Dr. Dennis Wittmer, Chair, Institutional Review Board for the Protection of Human Subjects, at (303) 871-2431 or Sylk Sotto-Santiago at (303) 871-4052.

The results of the final study will be available online at http://www.du.edu/~mearp/SAM. Upon graduation, I will update the webpage with a direct link to the published dissertation. I intend that the results of this study will benefit future generations of statistics students by providing a review of the factors that correlate with students suffering statistics anxiety.

Please sign below if you are willing to participate in a cognitive interview. Your participation is completely voluntary and will in no way affect your grade in your following statistics course. If you choose to participate, I will respond by scheduling a time for us to sit down either in person or over the phone to discuss the survey.

Thank you for your time and support!

Morgan S. Earp
(720) 217-5824

I understand that there are two exceptions to the promise of confidentiality. If information is revealed concerning suicide, homicide or child abuse and neglect, it is required by law that this be reported to the proper authorities. In addition, should any information contained in this study be subject of a court order or lawful subpoena, the University of Denver might not be able to avoid compliance with the order or subpoena.

I have read and understood the foregoing descriptions of the Development and Validation of the Statistics Anxiety Measure. I have asked for and received a satisfactory explanation of any language that I did not fully understand. I agree to participate in this study, and I understand that I may withdraw my consent at any time. I have received a copy of this consent form.

____
Signature Date

APPENDIX C

Item Pool for Statistics Anxiety Measure Study

Anxiety Factor

II. Please rate the following activities in terms of how anxious they make you feel.

	Not Anxious	Slightly Anxious	Anxious	Very Anxious
1. Taking statistics				
2. Taking a class that involves math				
3. Taking tests in this class				
4. Taking tests that involve mathematical computing				
5. Taking tests in general				
6. Interpreting and solving word problems				
7. Interpreting statistics				
8. Working with numbers				
9. Translating Greek symbols used in statistics				
10. Using a scientific calculator to calculate statistics				
11. Using a computer to calculate statistics				
12. Solving mathematical equations				
13. Using *z* or *t* tables				
14. Formulating and testing hypotheses				
15. Recalling mathematical principles				
16. Using the bell curve				
17. Reading statistical studies				
18. Calculating probabilities				
19. Interpreting mathematical solutions				
20. Explaining your statistical findings				

Fearful Behavior Factor

III. Please rate the following statements in terms of how much you agree or disagree with them.

	Strongly Disagree	Disagree	Agree	Strongly Agree
1. I have trouble staying mentally organized	○	○	○	○
2. I worry extensively about school	○	○	○	○
3. I worry extensively about taking this class	○	○	○	○
4. I fear asking the instructor for help	○	○	○	○
5. I am afraid of my statistics instructor	○	○	○	○
6. Taking this class makes me tense	○	○	○	○
7. I worry about doing well in this class	○	○	○	○
8. I second guess my ability to do well in this class	○	○	○	○
9. I expect I will be more hesitant to participate in this class	○	○	○	○
10. I am more stressed out this quarter as a result of having to take this class	○	○	○	○
11. I am afraid to ask for help in this class	○	○	○	○
12. I have spent more time worrying about keeping my GPA up this quarter than usual	○	○	○	○
13. I am worried this class will bring down my GPA	○	○	○	○
14. I had a hard time relaxing over break knowing I had to take statistics this quarter	○	○	○	○
15. I've been worried ever since I was informed this class was a requirement for my degree	○	○	○	○
16. I've avoided taking this class as long as possible	○	○	○	○
17. I was hesitant to register for this class	○	○	○	○
18. If there was a way I could avoid taking this class I would	○	○	○	○
19. I am so worried about doing well in this class that it makes it hard to concentrate	○	○	○	○
20. I am afraid of instructors in general	○	○	○	○

Attitude Factor

IV. Please rate the following statements in terms of how much you agree or disagree with them.

	Strongly Disagree	Disagree	Agree	Strongly Agree
1. I dislike math	○	○	○	○
2. I dislike working with numbers	○	○	○	○
3. I dislike solving word problems	○	○	○	○
4. It is pointless for me to take this class	○	○	○	○
5. This is my least exciting class	○	○	○	○
6. I expect this class to be boring	○	○	○	○
7. I am only taking this class because it is required	○	○	○	○
8. Taking this class will have little impact on my life	○	○	○	○
9. I do not expect to enjoy this class	○	○	○	○
10. My ability to calculate statistics will not affect my chances of getting a job	○	○	○	○
11. I will never use what I learn in this class again	○	○	○	○
12. I do not understand why I am required to take this class	○	○	○	○
13. Math is my least favorite subject	○	○	○	○
14. I'd rather read and write than do arithmetic	○	○	○	○
15. Statistics has nothing to do with my major	○	○	○	○
16. I usually skim over the reported statistics in the required reading for my courses	○	○	○	○
17. There is no room to be creative in statistics	○	○	○	○
18. I'd rather analyze a novel than numerical data	○	○	○	○
19. Statistics are just a scientific way of stating the obvious	○	○	○	○
20. I dislike calculating percentages	○	○	○	○

Expectation Factor

V. Please rate the following statements in terms of how much you agree or disagree with them.

	Strongly Disagree	Disagree	Agree	Strongly Agree
1. My parents and/or peers have unrealistically high expectations regarding my success in math related courses				
2. I expect to be punished if I fail to meet the demands of mathematical solving situations in this class				
3. My math reasoning ability is low				
4. I feel helpless when it comes to solving math problems				
5. I lack academic self confidence				
6. I am expected to do well in statistics				
7. People expect learning statistics to come naturally to me				
8. I have a hard time processing word problems				
9. Most people do well in statistics				
10. All my friends did well in statistics				
11. My mother or father did well in math related subjects				
12. My major involves working with numbers				
13. Failing this class could prevent me from graduating				
14. I struggle with mathematical reasoning				
15. I confuse the order of the steps I am supposed to follow when solving mathematical equations				
16. My advisor will be disappointed if I fail to do well in this class				
17. I am not good at math				
18. Math is the subject where I have the least amount of confidence				
19. My parents expect me to do well in this class				
20. I am afraid I'll let people down if I don't do well in this class				

History & Self-Concept Factor

VI. Please rate the following statements in terms of how much you agree or disagree with them.

	Strongly Disagree	Disagree	Agree	Strongly Agree
1. I have low self-esteem when it comes to math				
2. My math scores on the ACT/SAT were below average				
3. I have a history of failure in situations involving math				
4. I failed math in elementary and/or high school				
5. I am taking the minimum number of math courses required for my degree				
6. People expect me to do poorly in this class				
7. The quality of my math classes in the past was poor				
8. I question my instructor's ability to teach statistics				
9. I have disliked my math/statistics instructors in the past				
10. My instructor does not seem enthused to be teaching statistics				
11. Based on past experience, I expect the material covered in this class and the exams to be difficult				
12. This is my first statistics class ever				
13. My math and/or computer skills are lacking				
14. I lack motivation to learn or continue learning statistics				
15. I doubt my ability to learn statistics in a lecture environment				
16. I have never enjoyed classes that involve math				
17. My least enjoyable experiences in school involved math				
18. I've never enjoyed working with numbers				
19. I've struggled to follow the material covered in statistics classes in the past				
20. Based on past experiences I have little confidence in my ability to do arithmetic				

Performance Factor

VII. Please rate how well you expect to perform in the following areas.

	Above Average	Average	Below Average	Fail
1. Class exercises				
2. Homework				
3. Quizzes				
4. Exams				
5. Projects				
6. Reading assignments				
7. Solving mathematical equations				
8. Interpreting word problems				
9. Solving equations by hand				
10. Solving equations using the calculator				
11. Solving equations using the computer				
12. Interpreting my answers				
13. Following textbook examples				
14. Explaining my answers				
15. Making accurate conclusions based on statistical findings				
16. Teaching my peers				
17. Learning from my peers				
18. Testing hypotheses				
19. Selecting the appropriate statistical test				
20. Developing appropriate methodology to test a given hypothesis				

APPENDIX D

Class Participation Request Letter

Dear **[insert instructor's name]**,

Greetings! My name is Morgan Earp. I am a doctoral candidate in the College of Education's Quantitative Research Methods program. I am developing and validating a measure of statistics anxiety in partial fulfillment of the degree, doctorate of philosophy. As an undergraduate transfer student in 1999, I was required, as were all students, to take introduction to statistics. As a graduate student in the Graduate School of International Studies, I was required to take statistical methods. It was through these experiences that my interest in the topic of statistics anxiety began. Throughout my college experience I struggled with every class, that is until I discovered statistics. What intrigued me is that for the majority of my peers, the situation was opposite mine. Those who were fully confident of their academic abilities suddenly doubted their ability to succeed when it came time to take statistics. As I am sure you are aware the general public shares a distaste for this subject and thus many of our students hold negative views and/or doubt their ability to succeed before our class even starts.

I have spent the last four years researching the factors that comprise statistics anxiety. I am now moving into the final phase of my research at DU, and I am looking for your approval to administer my measure of statistics anxiety in your **[insert name of class]** on **[insert date of first day of class]** along with a brief demographic survey. I expect the Statistics Anxiety Measure (Earp, 2006) will take no more than five minutes for students to complete. I am also hoping to administer cross validation measures [the Survey of Attitudes Toward Statistics (SATS: Schau et al., 1995) and the Statistical Anxiety Rating Scale (STARS: Cruise & Wilkins, 1980)] in a limited number of classes with a total time commitment of no more than about 15 minutes.

All students will be informed both verbally and in writing that their participation is completely voluntary and that their decision to participate will in no way affect their final grade in your course. All volunteers will be rewarded with assorted candy upon finishing the questionnaire(s). The questionnaires are completely anonymous and will be placed directly by the student into a locked secure box provided by the researcher. The items appearing on the questionnaire(s) and the methodology for this study have been reviewed and approved by both my dissertation committee (Quantitative Research Methods Program Chair, Dr. Kathy Green; Associate Professor of Curriculum and Instruction, Dr. Nick Cutforth; and Graduate School of International Studies Director of Quantitative Studies and Senior Lecturer of the Department of Statistics and Operations Technology, Terry Dalton) and the Institutional Review Board at the University of Denver.

The results of this study will be available online at http://www.du.edu/~mearp/SAM. Upon graduation, the researcher will update the webpage with a direct link to the published dissertation. I intend that the results of this study will benefit you the instructor by providing a review of the factors that correlate with our students suffering statistics anxiety.

Please respond to this email and let me know if you are willing to let me administer my measure in your **(insert class name)** class on **(insert class date, day, and time).**

Thank you for your time and support! I look forward to hearing from you.

Morgan S. Earp

(720) 217-5824

APPENDIX E

Student Data Collection Information Sheet for Statistics Anxiety Measure Study

Dear Student:

This is to request your participation in a research study to explore attitudes toward statistics. I am asking you to respond to questions that ask about your attitude toward statistics along with some demographic items and your perception of how you expect to do in this course. This study is being conducted to fulfill the requirements of a doctoral dissertation. One benefit from participating in the study is that you can contribute to the improvement of future statistics curriculum and courses. **Your participation in this study is completely voluntary. Your decision to participate, not to participate, or to withdraw from the study, will not affect any grade in this course. If you are willing participate, please complete the attached questionnaire(s) and then place it in the opening of the locked secure box held by the researcher or the researcher's assistant.** The information you provide is completely anonymous. Only the researcher will see the completed questionnaire(s). Your name should not be written on the survey. Names will not be used in any reports of this study. There are no risks to participating that I am aware of, but participating will require about 20 minutes of your time. The questionnaires are coded with a number that will be used to identify the questionnaire and link it to the other questionnaires, for those participants asked to complete more than one.

Please keep the attached card and if you have any questions, please contact the researcher, Morgan Earp using the following email address: mearp@du.edu.

This study is being conducted by myself, Morgan Earp, with supervision by Dr. Kathy Green, College of Education, 303-871-2490, kgreen@du.edu, whom you may contact if you wish further information or if you have any concerns about the project. If you have any concerns or complaints about how you were treated during the process, please contact Dr. Dennis Wittmer, Chair, Institutional Review Board for the Protection of Human Subjects, at (303) 871-2431 or Sylk Sotto-Santiago at (303) 871-4052.

I am glad to share to the results of the study with you the participants. Please visit http://www.du.edu/~mearp/SAM for updated results and summaries of the study. Once the study is complete, a link to the published dissertation will be posted.

Sincerely,

Morgan S. Earp

APPENDIX F

Pilot Version of the Statistics Anxiety Measure

Date: ___/___/______
Time: ____:____ AM / PM
Course Number: ______________
Instructor: ______________

The SAM

I. Please answer the following demographic questions.

1. Gender:

   Female ☐ Male ☐

2. Race/Ethnicity:

   African American ☐ Asian/Pacific Islander ☐ Middle Eastern ☐
   Native American ☐ Caucasian/White ☐ None of the Above ____

3. Age ____
4. Major ____
5. Minor ____
6. Degree

   B.A. ☐ B.S. ☐ B.S.B.A. ☐ B.F.A. ☐
   M.A. ☐ M.S. ☐ M.B.A. ☐ M.F.A. ☐
   Ph.D. ☐ Psy.D. ☐ Ed.D. ☐

1 *Please turn the page over to continue*

II. Please rate the following activities in terms of how anxious they make you feel.

	Not Anxious	Slightly Anxious	Anxious	Very Anxious
1. Calculating probabilities	○	○	○	○
2. Taking statistics	○	○	○	○
3. Developing conclusions based on mathematical solutions	○	○	○	○
4. Solving mathematical equations	○	○	○	○
5. Reading statistical studies	○	○	○	○
6. Taking a class that involves math	○	○	○	○
7. Taking tests in this class	○	○	○	○
8. Interpreting statistics	○	○	○	○
9. Explaining your statistical findings	○	○	○	○
10. Formulating and testing hypotheses	○	○	○	○

III. Please rate the following statements in terms of how much you agree or disagree with them.

	Strongly Disagree	Disagree	Agree	Strongly Agree
1. If there was a way I could avoid taking this class I would	○	○	○	○
2. I am worried about taking statistics	○	○	○	○
3. I don't see myself participating in this class	○	○	○	○
4. I am worried this class will bring down my GPA	○	○	○	○
5. I worry about doing well in this class	○	○	○	○
6. Taking this class stresses me out	○	○	○	○
7. I've avoided taking this class as long as possible	○	○	○	○
8. I am afraid to ask for help in this class	○	○	○	○
9. I've been worried ever since I was informed this class was a requirement for my degree	○	○	○	○
10. I was hesitant to register for this class	○	○	○	○

Please continue on the next page

IV. Please rate the following statements in terms of how much you agree or disagree with them.

	Strongly Disagree	Disagree	Agree	Strongly Agree
1. I am only taking this class because it is required				
2. I do not expect to enjoy this class				
3. I expect this class to be boring				
4. Math is my least favorite subject				
5. I dislike math				
6. I dislike working with numbers				
7. I will never use what I learn in this class again				
8. There is no room to be creative in statistics				
9. Taking this class will have little impact on my life				
10. My ability to calculate statistics will not affect my chances of getting a job in my chosen field				

V. Please rate the following statements in terms of how much you agree or disagree with them.

	Strongly Disagree	Disagree	Agree	Strongly Agree
1. Math is the subject where I have the least amount of confidence				
2. I am not good at math				
3. My math reasoning ability is low				
4. My major involves working with numbers				
5. My mother or father did well in math related subjects				
6. I lack academic self confidence				
7. My parents expect me to do well in this class				
8. Most people do well in statistics				
9. I am expected to do well in statistics				
10. I feel helpless when it comes to solving math problems				

3

Please turn the page over to continue

VI. Please rate the following statements in terms of how much you agree or disagree with them.

	Strongly Disagree	Disagree	Agree	Strongly Agree
1. I have never enjoyed classes that involve math				
2. I am taking the minimum number of math courses required for my degree				
3. I lack motivation to learn or continue learning statistics				
4. Based on past experience, I expect the material covered in this class and the exams to be difficult				
5. This is my first statistics class ever				
6. My least enjoyable experiences in school involved math				
7. I've never enjoyed working with numbers				
8. I've struggled to follow the material covered in statistics classes in the past				
9. I doubt my ability to learn statistics in a lecture environment				
10. I have low self-esteem when it comes to math				

VII. Please rate how well you expect to perform in the following areas.

	Above Average	Average	Below Average	Fail
1. Explaining my answers				
2. Quizzes				
3. Projects				
4. Solving mathematical equations				
5. Solving equations by hand				
6. Interpreting my answers				
7. Making accurate conclusions based on statistical findings				
8. Exams				
9. Solving equations using the calculator/computer				
10. Developing appropriate methodology to test a given hypothesis				

☺ *Thank you for completing the SAM!* ☺

APPENDIX G

Final Administration Version of the Statistics Anxiety Measure

Date: ____/____/_______
Time: ____:____ AM / PM
Course Number: ______________
Instructor: __________________

The SAM

I. Please answer the following demographic questions.

1. Gender:

Female ○ Male ○

2. Race/Ethnicity:

African American ○ Asian/Pacific Islander ○ Latin American ○

Middle Eastern ○ Native American ○ Caucasian/White ○

None of the Above (Please specify) __________

3. Age ______

4. Major __________

5. Minor __________

6. Type of Degree you are currently seeking
(Please select more than one if you are currently working on a dual degree)

B.A. ○ B.S. ○ B.S.B.A. ○ B.F.A. ○ B.A./M.A. (4 +1) ○

M.A. ○ M.S. ○ M.B.A. ○ M.F.A. ○

Ph.D. ○ Psy.D. ○ Ed.D. ○ J.D. ○

1

Please turn the page over to begin

II. Please rate the following activities in terms of how anxious they make you feel.

	Not Anxious	Slightly Anxious	Anxious	Very Anxious
1. Taking tests in this class				
2. Explaining your statistical findings				
3. Formulating and testing hypotheses				
4. Interpreting statistics				
5. Taking statistics				
6. Calculating probabilities				
7. Developing conclusions based on mathematical solutions				
8. Reading statistical studies				

Please rate the following statements in terms of how much you agree or disagree with them.

	Strongly Disagree	Disagree	Agree	Strongly Agree
9. I am worried about taking statistics				
10. Based on past experience, I expect the material covered in this class and the exams to be difficult				
11. I worry about doing well in this class				
12. I've struggled to follow the material covered in statistics classes in the past				

Please continue on the next page

III. Please rate the following statements in terms of how much you agree or disagree with them.

	Strongly Disagree	Disagree	Agree	Strongly Agree
1. If there was a way I could avoid taking this class I would	○	○	○	○
2. I do not expect to enjoy this class	○	○	○	○
3. I expect this class to be boring	○	○	○	○
4. I am only taking this class because it is required	○	○	○	○
5. I will never use what I learn in this class again	○	○	○	○
6. My ability to calculate statistics will not affect my chances of getting a job in my chosen field	○	○	○	○
7. I lack motivation to learn or continue learning statistics	○	○	○	○
8. Taking this class will have little impact on my life	○	○	○	○
9. There is no room to be creative in statistics	○	○	○	○

IV. Please rate the following statements in terms of how much you agree or disagree with them.

	Strongly Disagree	Disagree	Agree	Strongly Agree
1. I was hesitant to register for this class	○	○	○	○
2. I am afraid to ask for help in this class	○	○	○	○
3. I've been worried ever since I was informed this class was a requirement for my degree	○	○	○	○
4. I've avoided taking this class as long as possible	○	○	○	○

Please turn the page over to continue

V. Please rate the following statements in terms of how much you agree or disagree with them.

	Strongly Disagree	Disagree	Agree	Strongly Agree
1. My math reasoning ability is low				
2. Math is my least favorite subject				
3. I dislike math				
4. I've never enjoyed working with numbers				
5. I am not good at math				
6. I have never enjoyed classes that involve math				
7. I dislike working with numbers				
8. I have low self-esteem when it comes to math				
9. Math is the subject where I have the least amount of confidence				
10. My least enjoyable experiences in school involved math				

VI. Please rate how well you expect to perform in the following areas.

	Above Average	Average	Below Average	Fail
1. Developing appropriate methodology to test a given hypothesis				
2. Solving equations using the calculator/computer				
3. Projects				
4. Making accurate conclusions based on statistical findings				
5. Exams				
6. Explaining my answers				
7. Quizzes				
9. Interpreting my answers				

4 ☺ *Thank you for completing the SAM!* ☺