A CORRELATIONAL ANALYSIS OF TEST ANXIETY AND RESPONSE TIME ON A COMPUTERIZED ADAPTIVE MATH TEST AMONG SEVENTH GRADE STUDENTS BY GENDER

by

Michael Wayde Brom
Liberty University

A Dissertation Presented in Partial Fulfillment Of the Requirements for the Degree Doctor of Education

Liberty University
2016
A CORRELATIONAL ANALYSIS OF TEST ANXIETY AND RESPONSE TIME ON A
COMPUTERIZED ADAPTIVE MATH TEST AMONG SEVENTH GRADE STUDENTS
BY GENDER

by Michael Wayde Brom

A Dissertation Presented in Partial Fulfillment
Of the Requirements for the Degree
Doctor of Education

Liberty University, Lynchburg, VA
2016

APPROVED BY:
Sarah Hutter, Ed.D., Committee Chair
Glenn Holzman, Ph.D., Committee Member
Lori Benton, Ph.D., Committee Member
Scott Watson, Ph.D., Associate Dean, Advanced Programs
ABSTRACT

Studies have shown that test anxiety has become more prevalent since the adoption of the No Child Left Behind Act in 2001 and that test anxiety negatively affects student achievement. Early research viewed test anxiety as being a unidimensional construct; however, recent research has purported that test anxiety is a multidimensional construct. Consequently, test anxiety being viewed as multidimensional has changed how test anxiety is measured as evidenced by the development of multidimensional test anxiety measurement scales. Literature indicates that there is a gap in the study of K-12 test anxiety and response time data on assessments. A nonexperimental correlational research design was used for this study. The purpose of this study was to examine the relationships between multidimensional test anxiety measures and the response time of students taking a computerized adaptive math assessment, the Measures of Academic Progress® (MAP®) by the Northwest Evaluation Association™ (NWEA™), while controlling for gender. Participants for the study were 215 seventh grade math students at a school in the West. The test anxiety of participants was measured using the Children’s Test Anxiety Scale (CTAS). The response time data was captured by the NWEA™ MAP® math tests. Spearman correlations were used to examine the relationship between test anxiety and response time. No statistically significant relationships were found. Recommendations for future research are to examine relationships between test anxiety and RTE using different normative thresholds (NT) and utilizing other multidimensional test anxiety scales as they become commercially available.

Keywords: computerized adaptive testing, deficit model, expectancy theory, interference model, response time, test anxiety
Dedication

For Mom and Dad, as well as my three sons, Cody, Dominic, and Noah. Your ongoing support, patience, encouragement, and the many sacrifices you have each made as I worked on my degree sustained me. All of you contributed to this tremendous accomplishment, and I gladly share it with you. You are my heroes and are a tremendous blessing to me. You helped see me through to the very end, which has resulted in me achieving my lifelong goal of earning my doctorate. I love you all very much.
Acknowledgments

I thank my Lord and Savior, Jesus Christ, for leading me to Liberty University and for giving me the strength and perseverance to finish what I started in earning this degree. I also wish to thank God for placing the right people in my life at the right time, both prior to and during my doctoral studies, to prepare me for and assist me in earning my doctorate.

Thank you to my mother and former high school English teacher, Mrs. (Joyce) Brom, for her countless hours of proofreading and for helping me develop the writing skills necessary to produce a dissertation I could be proud of. I can never thank you enough, Mom, for your willingness to provide me with constructive feedback at a moment’s notice and for your keen insights into my strengths and weaknesses as a writer. You had a profound influence in the writing quality of my dissertation. Thank you, Dad, for your consistent “you can do it!” reminders, especially during my moments of weakness and self-doubt. You were a calming presence during those times when I felt like giving up.

Thank you to Northwest Evaluation Association™ (NWEA™) for their generous support and assistance with the response time effort data I needed for my research. This study would not have been possible without the participants’ response time data NWEA™ provided to me.

I owe a debt of gratitude to my dissertation committee. I wish to thank my chair, Dr. Sarah Hutter, for her willingness to serve as my chair, for her guidance, support, and invaluable constructive feedback, as well as for the ongoing, timely affirmation of my work which served to motivate and inspire me. Thank you for keeping me grounded throughout the doctoral process, Dr. Hutter. Thank you as well to Dr. Glenn Holzman and Dr. Lori Benton for your willingness to serve on my committee. Your time, insights, and thoughtful feedback are greatly appreciated. I am so blessed to have had such a wonderful committee to assist me in my doctoral journey.
# Table of Contents

ABSTRACT ............................................................................................................................... 3

Dedication ............................................................................................................................... 4

Acknowledgments .................................................................................................................... 5

List of Figures .......................................................................................................................... 10

List of Abbreviations ............................................................................................................. 11

CHAPTER ONE: INTRODUCTION ......................................................................................... 12

Background .............................................................................................................................. 12

Problem Statement .................................................................................................................. 19

Purpose Statement ................................................................................................................... 20

Significance of the Study ......................................................................................................... 21

Research Questions ................................................................................................................. 23

Null Hypotheses ...................................................................................................................... 24

Definitions .............................................................................................................................. 27

CHAPTER TWO: LITERATURE REVIEW .................................................................................. 30

Introduction .............................................................................................................................. 30

Theoretical Frameworks ......................................................................................................... 35

Social Learning Theory ........................................................................................................... 35

Expectancy Theory .................................................................................................................. 40

Related Literature ................................................................................................................... 45
Models of Test Anxiety ................................................................. 45
Test Anxiety Models Summary ......................................................... 50
Effects of Test Anxiety on Student Groups........................................... 51
Online Assessments and Paper-Based Assessments.................................. 55
Measures of Academic Progress® (MAP®) ................................................ 58
Response Time Effort (RTE) .................................................................. 59
Identifying Students with Test Anxiety .................................................... 62
Test Anxiety Scales ............................................................................... 67

Summary .......................................................................................... 71

CHAPTER THREE: METHODS ............................................................... 73

Design ............................................................................................... 73
Research Questions ............................................................................ 75
Null Hypotheses ................................................................................ 76
Participants and Setting ....................................................................... 79
Instrumentation .................................................................................. 81
Procedures .......................................................................................... 84
Data Analysis ...................................................................................... 86

CHAPTER FOUR: FINDINGS ................................................................. 90

Research Questions ............................................................................ 90
Null Hypotheses ................................................................................ 90
Descriptive Statistics ................................................................. 93

Results ...................................................................................... 94

Null Hypothesis One ................................................................. 94

Null Hypothesis Two ................................................................. 95

Null Hypothesis Three ............................................................. 95

Null Hypothesis Four ............................................................... 95

Null Hypothesis Five ............................................................... 96

Null Hypothesis Six ................................................................. 96

Null Hypothesis Seven .............................................................. 96

Null Hypothesis Eight ............................................................... 97

Null Hypothesis Nine ............................................................... 97

Null Hypothesis Ten ................................................................. 98

Null Hypothesis Eleven ............................................................ 98

Null Hypothesis Twelve .......................................................... 98

Null Hypothesis Thirteen .......................................................... 99

Null Hypothesis Fourteen ......................................................... 99

Null Hypothesis Fifteen ............................................................ 100

Null Hypothesis Sixteen ........................................................... 100

CHAPTER FIVE: DISCUSSION, CONCLUSIONS, AND RECOMMENDATIONS .......... 101

Discussion .................................................................................. 101
Research Question One........................................................................................................ 104

Research Question Two........................................................................................................ 105

Conclusions.......................................................................................................................... 106

Implications.......................................................................................................................... 107

Limitations ............................................................................................................................ 109

Recommendations for Future Research ............................................................................. 111

REFERENCES..................................................................................................................... 116

APPENDICES ...................................................................................................................... 132

Appendix A: Children’s Test Anxiety Scale ................................................................. 132

Appendix B: Original Children’s Test Anxiety Scale.................................................... 135

Appendix C: Permission Letter to Reproduce and Publish the Children’s Test Anxiety
Scale ...................................................................................................................................... 138

Appendix D: Institutional Review Board Approval Letter ......................................... 139

Appendix E: Participant Consent Form ........................................................................... 140

Appendix F: Assent Form ................................................................................................. 143

Appendix G: Recruitment Letter ...................................................................................... 144

Appendix H: Institutional Review Board Change in Protocol Letter ............................ 145
List of Figures

Figure 1. Test Anxiety Response Time Correlations. ............................................................ 75
List of Abbreviations

Aggregate Response Time (ART)
Children’s Test Anxiety Scale (CTAS)
Cognitive Behavior Therapy (CBT)
Measures of Academic Progress® (MAP®)
No Child Left Behind (NCLB)
Northwest Evaluation Association™ (NWEA™)
Normative Threshold (NT)
Partnership for Assessment of Readiness for College and Careers (PARCC)
Response Time Effort (RTE)
Rasch Unit (RIT)
Smarter Balanced Assessment Consortium (SBAC)
Stereotype Threat (ST)
Test Anxiety Inventory (TAI)
Test Anxiety Inventory for Children and Adolescents (TAICA)
Test Anxiety Questionnaire (TAQ)
Test Anxiety Scale (TAS)
Test Anxiety Scale for Children (TASC)
CHAPTER ONE: INTRODUCTION

Background

Test anxiety steadily increased among K-12 students through the 1980s and 1990s; the increase in test anxiety among K-12 students has continued since the passage of the No Child Left Behind (NCLB) Act (Cassady, 2010; Huberty, 2010; Segool, Carlson, Goforth, von der Embse, & Barterian, 2013; von der Embse & Hasson, 2012; von der Embse, Mata, Segool, & Scott, 2014), which subsequently resulted in an increase in testing in reading and mathematics (No Child Left Behind [NCLB], 2002). As a result the amount of test anxiety experienced by students is expected to increase with the increase in the amount of testing as well as the increase in testing requirements as mandated by NCLB (Wren & Benson, 2004). Methia (2004) estimated that 33% of all elementary and secondary students experienced some measure of test anxiety. More recently others estimated that test anxiety affected between 25% and 40% of students (Cassady, 2010; Huberty, 2010). With more students taking high-stakes assessments and students experiencing test anxiety at younger ages, there was an expectation that test anxiety prevalence would continue to increase (von der Embse & Hasson, 2012).

In an attempt to better identify not only the prevalence but also the specific dimensions of test anxiety, recent test anxiety research strived to refine test anxiety dimensions and, thus, develop refined test anxiety surveys (Lowe et al., 2008; Unruh & Lowe, 2010; Wren & Benson, 2004). Furthermore, the improvements in technology and computer-based assessments have provided practitioners with data typically unavailable or difficult to glean from paper-based assessments. One particular benefit of computer-based assessments is the ability to measure student effort in terms of response time, which is the aggregate time students spend taking an assessment. Given the dynamic nature of student effort throughout any given assessment,
response time research has focused not only on aggregate time but also response time effort (RTE), a summary measure of effort at the item level based on a response time threshold (Wise & Kong, 2005). Much of the RTE research has focused on student engagement, time of testing, motivation, and test validity. Relationships between response time, both aggregate response time and response time effort (RTE), and test anxiety in K-12 education, which was the focus of this study, had not yet been examined. The sections that follow highlight key background information about the study, the problem statement associated with the study, the purpose and significance of the study, the research questions and associated research hypotheses, as well as the assumptions and limitations of the study.

Over the past 60 years test anxiety evolved from a unidimensional construct to a multidimensional construct (Alpert & Haber, 1960; Hembree, 1988; Hodapp & Benson, 1997; Liebert & Morris, 1967; Mandler & Sarason, 1952; Sarason, 1984; Wine, 1971; Zeidner, 1998). Early researchers focused on the test anxiety construct itself in an effort to define, measure, and determine the impact test anxiety had on students and student achievement. The passage of the NCLB Act in 2001 brought about a significant expansion in the number of mandated state assessments taken by students in grades three and above. It also generated a renewal of interest in understanding test anxiety, its impact on student achievement, and a renewal of efforts in reducing test anxiety.

A report showed that test anxiety levels increased for up to 25% of students when students self-reported test anxiety for classroom assessments and self-reported test anxiety for NCLB-mandated state assessments (Segool et al., 2013). Sixty percent of the students in the Segool et al. (2013) study showed stable test anxiety levels regardless of testing conditions. Thus, 40% of students had a change in test anxiety levels where 15% of students had less test
anxiety about the NCLB-mandated state assessment than about the classroom assessment. This may have potentially resulted from differences between the environment of a state assessment setting and the environment of a classroom assessment setting.

Teacher interaction with students was found to be a factor in test anxiety. Without receiving positive reinforcement from teachers, the learning motivation and academic self-concept of students may decrease while test anxiety may increase (Urhanhe, Chao, Florineth, Luttenberger, & Paechter, 2011). The variance in student achievement scores attributed to students’ test anxiety levels ranged from four percent to as much as 15% (Cassady & Johnson, 2002; Putwain, 2008c; Seipp, 1991; von der Embse & Hasson, 2012). The validity of the very test scores themselves was threatened by construct-irrelevant variance, which were uncontrolled variables that affected test outcomes (Haladyna & Downing, 2004). This included uncontrolled variables, such as item bias, guessing, testwiseness, and poorly written test questions. The validity of computerized adaptive tests can also be influenced by both students and teachers. Student motivation and teachers who do not place the same emphasis on both fall and spring administrations of a computerized adaptive test may also have had a negative impact on individual score validity (ISV) (Wise, Ma, Cronin, & Theaker, 2013).

In an era of increased accountability for school districts, schools, and teachers, the range of construct-irrelevant variance may be significant not only for students but also for educators. Student scores on NCLB-mandated state assessments are now being factored into teacher evaluations, tenure, and even pay in some states, such as Florida, Georgia, and Tennessee, which were three of the first 12 states to receive grants through the federal Race to the Top (RTT) initiative (Boser, 2012). One of the eight criteria in the RTT application was, in fact, that “teacher evaluation results be used to guide decisions about compensation (including annual
salary increases or performance-based compensation)” (Hallgren, James-Burdumy, & Perez-Johnson, 2014, p. 5). Among 49 states and the District of Columbia, evaluation results were used to determine career advancement (one state), annual salary increases (six states), and performance-based compensation (five states) (Hallgren et al., 2014). To ensure valid assessment results, confounding variables should be taken into account when interpreting achievement scores.

Test anxiety was considered by some to be one of the most disruptive factors in test performance (Cizek & Burg, 2006). Many factors, including test anxiety, may have potentially interfered with the accuracy of student achievement results on assessments (Putwain, 2008a), as in the case where discrepancies on NCLB-mandated state assessments were found in the relative proficiency levels between reading and mathematics, as well as the proficiency levels between grade levels within a content area. Such discrepancies resulted in part from assessments that were not equidifficult and from the inconsistent setting of the proficiency levels, or cut scores, to differentiate between students whose performance was categorized as being unsatisfactory, partially proficient, proficient, or advanced (Cronin, Dahlin, Adkins, & Kingsbury, 2007).

Previous research not only indicated that high-stakes testing situations increased test anxiety but also that time limitations increased test anxiety (Zeidner, 1998). Wilhelm and Schulze (2002) argued that when time limits for assessments were imposed on students, some students hurried through the assessment, which perhaps impacted student achievement. Thus, whenever students adhered to time limits for an assessment, students may have increased their test-taking speed at the expense of accuracy or they may have sacrificed accuracy in an attempt to complete the test within the time limits of the assessment (Klein Entink, Hornke, Kuhn, & Fox, 2009; Partchev, De Boeck, & Steyer, 2013). Zuriff (1997) suggested that students with test
anxiety be permitted to take tests without time limits. Nonetheless, Wise (2014) used time limits in an effort to standardize test administrations in order to avoid introducing construct-irrelevant variance which would potentially undermine test score validity.

Motivation and test anxiety relationships were found to have impacted assessment results (Wolf & Smith, 1995). While examining a synthesis of 15 test-taking motivation studies, Wise and DeMars (2005) found that the mean test performance of students identified as having low motivation was .58 standard deviations lower than the mean test performance of their more highly motivated peers. The effort students put forth on an assessment may then have been a necessary factor to consider when interpreting test results, as high levels of effort were more likely to accurately reflect the cognitive ability of students while low levels of effort make drawing inferences about the cognitive ability of students less reliable (Silm, Must, & Täht, 2013).

The aggregate response time, and even moreso the response time effort (RTE), of students are two measures of motivation. Furthermore, student motivation was found to have impacted test validity as evidenced by motivation filtering, which demonstrated that removing the test scores of unmotivated students improved the validity and reduced the variance of test scores, which also consequently improved the quality of proficiency estimates (Wise & DeMars, 2010). It would behoove researchers and practitioners to critically examine variables, such as RTE and test anxiety, which may “interfere with the authentic measurement of student achievement and school effectiveness” (von der Embse & Hasson, 2012, p. 180). Research about the relationship between RTE and test anxiety is lacking, which presented an opportunity for this study to address a gap in test anxiety literature.
The conditions surrounding testing, such as whether a test is a high-stakes or low-stakes assessment, may impact students’ motivation, effort, test anxiety, test validity, and, consequently, interpretation of students’ scores, particularly when students do not value the scores they receive on a given assessment, and which subsequently, influence test performance (Wolf & Smith, 1995). Differentiating between what constitutes a high-stakes and low-stakes assessment is problematic within education. While tests such as the NCLB-mandated state assessments are high-stakes for those administering the test because of the accountability factor, the lack of consequences for students taking those state assessments would classify those assessments as being low-stakes (Wise & DeMars, 2010). For the purpose of this study, the stakes of an assessment were determined by the consequences, or lack thereof, from the perspective of the test-taker – the student.

One presumption about student motivation on assessments was that all students put forth sufficient effort and that students’ scores were a true representation of student proficiency levels (Setzer, Wise, van den Heuval, & Ling, 2013; Wise & DeMars, 2005, 2010). A lack of sufficient effort could have a negative impact on test performance, which would result in a test score that underestimated a student’s actual proficiency level (Wise & DeMars, 2010). This is problematic, especially when assessment data is used for accountability purposes for the test giver, student placement, or for making instructional program decisions. Advances in technology and the evolution of computer-based assessments have made it easier to unobtrusively measure student motivation by examining aggregate response time and response time effort (RTE). Effort, therefore, could be measured both by the aggregate response time and by the RTE of students. Given the presence of confounding variables in test scores, it could be “complicated to get ‘pure’ test scores” (Silm et al., 2013, p. 433). Nevertheless, it is not only
crucial that test results are a true measure of student knowledge but also that students’ well-being is a focus of practitioners.

Two theories relevant to this study are Bandura’s (1986) social learning theory and Vroom’s (1964) expectancy theory. Bandura touted in Social Learning Theory that three things impacted individuals: a) a person’s biological and psychological characteristics, b) a person’s behavior, and c) the person’s environment. A key aspect of social learning theory is self-efficacy, the conviction one has in his or her ability to be successful, which was used to explain differences in student performance in conjunction with test anxiety (Bandura, 1986). Self-efficacy is impacted by both academic and social behaviors. Three environments impacted one’s self-efficacy: a) imposed, b) selected, and c) created (Bandura, 1997).

Vroom (1964) noted that a student’s expectancy of achieving a certain performance outcome influenced the effort a student expends in order to reach that outcome. An expectancy-value interaction exists such that what a student expects in terms of success of failure is also influenced by what the student values. Wolf and Smith (1995) found that students with the highest levels of both motivation and test anxiety performed comparably to students with the lowest levels of motivation and test anxiety, which would suggest that the “debilitating effects of high (test) anxiety did not override the benefits of high motivation” (p. 238). Low motivation and testing conditions were shown to influence students’ efforts on math test items as well as impact the validity of data and, consequently, the accuracy of estimating students’ true abilities (Putwain, 2008b). A student’s test anxiety is impacted in part by his or environment and the effort he or she puts forth on any given assessment is affected, in part, by what he or she expects to accomplish and what he or she values. Thus, the theoretical frameworks of Bandura’s (1986) social learning theory and Vroom’s (1964) expectancy theory were enjoined for this study.
Problem Statement

Since the passage of NCLB and the subsequent increase in testing in reading and mathematics, test anxiety among students increased and was expected to continue to increase (Wren & Benson, 2004). Debilitating test anxiety was estimated to negatively affect between 25% and 40% of elementary and secondary students (Cassady, 2010; Huberty, 2010; Methia, 2004). This indicated an increase in the percentage estimates of students experiencing debilitating test anxiety throughout the 1990s (Friedman & Bendas-Jacob, 1997; Zeidner, 1998). In addition, females, students with disabilities, and minority students were found to have higher levels of debilitating test anxiety (Putwain, 2007; Whitaker Sena, Lowe, & Lee, 2007).

Debilitating test anxiety was negatively correlated with student performance on assessments, especially in mathematics (von der Embse & Hasson, 2012). Nonetheless, the relationship between test anxiety and student achievement should not be assumed to be causal (Putwain, Conners, & Symes, 2010a). Such an assumption does not take into consideration other factors that may contribute to low student achievement for students who experience high levels of debilitating test anxiety. One such factor is student motivation as measured by the aggregate response time and also by the response time effort (RTE), which was a focus of this study. The gap in research that this study addressed was the relationship between test anxiety and response time. The problem is that test anxiety has become more widespread among K-12 students and has been shown to have a negative effect on student achievement, and student effort, as measured by RTE, has also been shown to have a negative relationship with student achievement (Silm et al., 2013; von der Embse & Hasson, 2012; Wise, 2014).
Purpose Statement

The purpose of this nonexperimental correlational study was to determine the relationships between students’ response time, at both the aggregate and per item level (response time effort), while taking a computerized adaptive test, Northwest Evaluation Association™ (NWEA™) Measures of Academic Progress® (MAP®) math test, and test anxiety, while controlling for gender for seventh grade math students at a school district in the West. The variables of interest were test anxiety, which was defined as the Total Test Anxiety as measured by the multidimensional test anxiety survey, the Children’s Test Anxiety Scale (CTAS), each of the CTAS subscales which were Thoughts, Off-Task Behaviors, and Autonomic Reactions; aggregate response time; and RTE, which was a measure of students’ motivation based on the proportion of students’ solution behaviors.

Thoughts includes the worry cognitions that students experience during testing; Off-Task Behaviors includes the nervous habits and distracting behaviors that can be observed of students during testing; and Autonomic Reactions includes the physiological arousal symptoms, such as perspiring and stomach problems students experience during testing (Wren & Benson, 2004). Solution behaviors were distinguished from rapid-guessing behaviors by a predetermined time threshold for each item (Wise & Kong, 2005). The control and intervening variable was gender. This study sought to address a gap in test anxiety literature, as very few studies have been conducted which examined relationships between test anxiety and student motivation as measured either by the aggregate response time or by the RTE on computerized adaptive tests (Segool et al., 2013).

The ever increasing prevalence of debilitating test anxiety in recent years, as well as the expansion of testing as a result of the passage of NCLB, has heightened awareness of and the
need to determine relationships of test anxiety with other factors in student achievement, whether it be performance, intervention strategies, or motivation. It was hypothesized that students who were unmotivated tend to answer test items too quickly (Wise & Kong, 2005). Even prior to the implementation of NCLB, The Standards for Educational and Psychological Testing (American Educational Research Association et al, 1999) recommended that the test-taking efforts of students be used in the interpretation of test scores.

Although the response time effort (RTE) was examined for a low-stakes assessment by isolating time spent on each individual item, insufficient research exists regarding the total amount of time and the RTE of students taking a computerized adaptive test as a correlation with test anxiety (Wise & Kong, 2005). Additional studies which examine student response time and RTE data will aid in identifying student perseverance as a possible factor to address as a potential test anxiety intervention strategy. Computer-based technology is needed, however, to provide accurate RTE data.

It should be noted that in the Wise and Kong (2005) study, low-stakes assessments were also defined as being assessments with no personal consequences for students. This clarification was essential as students in some states must pass state assessments in order to pass to the next grade level or to graduate, which Wise and Kong (2005) defined as being high-stakes assessments. In other states, however, there were no personal consequences for how students performed on state assessments, which classified those particular assessments as being low-stakes assessments.

**Significance of the Study**

The significance of this study was that it addressed a gap in research regarding the relationship of test anxiety with aggregate response time and with response time effort (RTE).
Previous RTE research focused on relationships between motivation and student performance (Wise & Kong, 2005). This study examined the relationships between test anxiety and student motivation, which was measured by both students’ aggregate response time as well as their RTE, demonstrated while taking a computerized adaptive assessment, the NWEA™ Measures of Academic Progress® (MAP®) math test. Furthermore, while very little research exists of test anxiety in the context of online assessments, including computerized adaptive tests, the test anxiety students experience in the classroom does not necessarily strongly translate to the test anxiety students experience when taking online assessments (Stowell, Allan, & Teoro, 2012). These response time-test anxiety relationships on the MAP® math test were also examined for the student subsample of gender.

Students across the United States are affected by test anxiety on quizzes, unit tests, semester tests, computerized adaptive tests, state assessments, and standardized tests. An abundance of replicated research demonstrated the negative impact debilitating test anxiety has had on student achievement scores (Hembree, 1988; Putwain et al., 2010a; Reeve & Bonaccio, 2008; Seipp, 1991; von der Embse & Hasson, 2012; Wolf & Smith, 1995). Thus, an inverse relationship was found between students’ test anxiety levels and their student achievement. Furthermore, negative correlations of test anxiety and academic performance indicators were determined to be between .22 and .26 (Cassady & Johnson, 2002; Putwain et al., 2010a). Of even greater importance beyond the impact test anxiety had on student achievement was the potentially negative impact test anxiety may have had on the well-being of students (Harpell & Andrews, 2013).
Research Questions

The relationship between aggregate response time and test anxiety, as well as the relationships between response time effort (RTE) and test anxiety, were examined in the context of an unspeeded cognitive computerized adaptive math assessment, the Northwest Evaluation Association’s™ (NWEA™) Measures of Academic Progress® (MAP®) math test. Potential relationships were examined between the RTE, which is equal to the proportion of a student’s responses categorized as solution behaviors as defined by an established normative threshold that students spend taking the MAP® math assessment, and test anxiety.

Previous response time effort research used a Common 3 threshold, an arbitrary time whereby solution behavior was defined as being when a student spent three or more seconds on a given test item (Wise, Kingsbury & Hauser, 2009; Wise, Ma, Kingsbury & Hauser, 2010). The RTE values for each participant in this study were calculated using a threshold which were determined by combining the principles of the Common 3 method and the normative threshold 10 (NT10) methods. For this study a normative threshold of 10% of the mean aggregate response time per test item of the participants was used, because response times available to the researcher for the MAP® math tests were rounded to the nearest second. The Children’s Test Anxiety Scale (CTAS) (Wren & Benson, 2004), a multidimensional test anxiety self-report survey, was used to measure test anxiety for this study. The multidimensional subscales of CTAS are Thoughts, Off-Task Behaviors, and Autonomic Reactions. Relationships between response time, both aggregate response time and RTE, and Total Test Anxiety, as well as each of the test anxiety subscales, were examined.

For the research questions, the four measures of test anxiety consisted of the three subscales of the Children’s Test Anxiety Scale (CTAS), Thoughts, Off-Task Behavior, and
Autonomic Reactions, as well as the Total Test Anxiety, which was a sum of the three CTAS subscales.

The research questions for this nonexperimental correlational study were as follows:

**RQ1:** Is there a relationship between any of the four measures of test anxiety as measured by the Children’s Test Anxiety Scale (CTAS) and the response time effort (RTE) of seventh grade math students taking the Measures of Academic Progress® (MAP®) math assessment after separately controlling for gender?

**RQ2:** Is there a relationship between any of the four measures of test anxiety as measured by the Children’s Test Anxiety Scale (CTAS) and the aggregate response time of seventh grade math students taking the Measures of Academic Progress® (MAP®) math assessment after separately controlling for gender?

**Null Hypotheses**

Sixteen hypotheses were examined as part of this study. The null hypotheses for the first research question, which pertained to relationships between the four measures of test anxiety as measured by the Children’s Test Anxiety Scale (CTAS) survey and the seventh grade math students’ response time effort (RTE) while taking the Northwest Evaluation Association’s™ Measures of Academic Progress® (MAP®) math assessment, in the study are as follows:

**H₀₁:** There is no statistically significant relationship between students’ Total Test Anxiety as measured by the Children’s Test Anxiety Scale (CTAS) and students’ response time effort (RTE) while taking the Measures of Academic Progress® (MAP®) math assessment.

**H₀₂:** There is no statistically significant relationship between students’ Total Test Anxiety as measured by the Children’s Test Anxiety Scale (CTAS) and students’ response
time effort (RTE) while taking the Measures of Academic Progress® (MAP®) math assessment after controlling for gender.

**H₀₃** There is no statistically significant relationship between the Thoughts subscale as measured by the Children’s Test Anxiety Scale (CTAS) and students’ response time effort (RTE) while taking the Measures of Academic Progress® (MAP®) math assessment.

**H₀₄** There is no statistically significant relationship between the Thoughts subscale as measured by the Children’s Test Anxiety Scale (CTAS) and students’ response time effort (RTE) while taking the Measures of Academic Progress® (MAP®) math assessment after controlling for gender.

**H₀₅** There is no statistically significant relationship between the Off-Task Behaviors subscale as measured by the Children’s Test Anxiety Scale (CTAS) and students’ response time effort (RTE) while taking the Measures of Academic Progress® (MAP®) math assessment.

**H₀₆** There is no statistically significant relationship between the Off-Task Behaviors subscale as measured by the Children’s Test Anxiety Scale (CTAS) and students’ response time effort (RTE) while taking the Measures of Academic Progress® (MAP®) math assessment after controlling for gender.

**H₀₇** There is no statistically significant relationship between the Autonomic Reactions subscale as measured by the Children’s Test Anxiety Scale (CTAS) and students’ response time effort (RTE) while taking the Measures of Academic Progress® (MAP®) math assessment.
\textbf{H}_08: There is no statistically significant relationship between the \textit{Autonomic Reactions} subscale as measured by the Children’s Test Anxiety Scale (CTAS) and students’ \textit{response time effort} (RTE) while taking the Measures of Academic Progress\textsuperscript{®} (MAP\textsuperscript{®}) math assessment after controlling for gender.

The null hypotheses for the second research question, which pertained to relationships between the students’ Total Test Anxiety as measured by the Children’s Test Anxiety Scale (CTAS) survey and the seventh grade math students’ aggregate response time while taking the Northwest Evaluation Association’s™ Measures of Academic Progress\textsuperscript{®} (MAP\textsuperscript{®}) math assessment, in the study are as follows:

\textbf{H}_09: There is no statistically significant relationship between students’ \textit{Total Test Anxiety} as measured by the Children’s Test Anxiety Scale (CTAS) and students’ \textit{aggregate response time} while taking the Measures of Academic Progress\textsuperscript{®} (MAP\textsuperscript{®}) math assessment.

\textbf{H}_010: There is no statistically significant relationship between students’ \textit{Total Test Anxiety} as measured by the Children’s Test Anxiety Scale (CTAS) and students’ \textit{aggregate response time} while taking the Measures of Academic Progress\textsuperscript{®} (MAP\textsuperscript{®}) math assessment after controlling for gender.

\textbf{H}_011: There is no statistically significant relationship between the \textit{Thoughts} subscale as measured by the Children’s Test Anxiety Scale (CTAS) and students’ \textit{aggregate response time} while taking the Measures of Academic Progress\textsuperscript{®} (MAP\textsuperscript{®}) math assessment.

\textbf{H}_012: There is no statistically significant relationship between the \textit{Thoughts} subscale as measured by the Children’s Test Anxiety Scale (CTAS) and students’ \textit{aggregate response time} while taking the Measures of Academic Progress\textsuperscript{®} (MAP\textsuperscript{®}) math assessment after controlling for gender.
response time while taking the Measures of Academic Progress\textsuperscript® (MAP\textsuperscript®) math assessment after controlling for gender.

**H\textsubscript{013}:** There is no statistically significant relationship between the Off-Task Behaviors subscale as measured by the Children’s Test Anxiety Scale (CTAS) and students’ aggregate response time while taking the Measures of Academic Progress\textsuperscript® (MAP\textsuperscript®) math assessment.

**H\textsubscript{014}:** There is no statistically significant relationship between the Off-Task Behaviors subscale as measured by the Children’s Test Anxiety Scale (CTAS) and students’ aggregate response time while taking the Measures of Academic Progress\textsuperscript® (MAP\textsuperscript®) math assessment after controlling for gender.

**H\textsubscript{015}:** There is no statistically significant relationship between the Autonomic Reactions subscale as measured by the Children’s Test Anxiety Scale (CTAS) and students’ aggregate response time while taking the Measures of Academic Progress\textsuperscript® (MAP\textsuperscript®) math assessment.

**H\textsubscript{016}:** There is no statistically significant relationship between the Autonomic Reactions subscale as measured by the Children’s Test Anxiety Scale (CTAS) and students’ aggregate response time while taking the Measures of Academic Progress\textsuperscript® (MAP\textsuperscript®) math assessment after controlling for gender.

**Definitions**

The terms that follow were utilized as part of this study.

1. **Unidimensional Test Anxiety** – test anxiety defined as having a single attribute; individuals either have or do not have test anxiety (Sarason, 1961).
2. **Multidimensional Test Anxiety** – test anxiety construct consisting of cognitive, emotional, behavioral and physiological components (Sarason, 1984).


4. **Thoughts** – this CTAS subscale includes the worry cognitions that students experience during testing (Wren & Benson, 2004).

5. **Off-Task Behaviors** – this CTAS subscale includes the nervous habits and distracting behaviors that can be observed of students during testing (Wren & Benson, 2004).

6. **Autonomic Reactions** – this CTAS subscale includes the physiological arousal symptoms, such as perspiring and stomach problems students experience during testing (Wren & Benson, 2004).

7. **Response Time Threshold** – the time, in terms of seconds, which differentiates between a student’s solution behavior and a student’s rapid-guessing behavior for each test item (Wise & Kong, 2005).

8. **Aggregate Response Time** – the total time, or duration, a student spends to complete a test.

9. **Response Time Effort (RTE)** – the proportion, represented as a decimal value, of test items for which a student exhibited solution behavior (Wise & Kong, 2005).

10. **Solution Behavior** – behavior that is classified as being effortful, as determined by a set response time threshold, when a student answers a given item (Setzer et al., 2013).
11. Rapid-Guessing Behavior – behavior that is classified as being noneffortful, as determined by a set response time threshold, when a student answers a given item (Setzer et al., 2013).

12. High-Stakes Testing – when there are consequences for test performance from the perspective of the test-taker (Setzer et al., 2013).

13. Low-Stakes Testing – when there is an absence of consequences for test performance from the perspective of the test-taker (Setzer et al., 2013).

14. Measurement Bias – two test-takers with the same cognitive ability but who differ in test anxiety will have different expected test scores (Wicherts & Scholten, 2010).

15. Deficit model – this model suggests that test anxiety does not cause lower test scores; the correlation between test anxiety and test scores exists because people with lower ability levels tend to have higher levels of test anxiety; thus, no measurement bias exists due to test anxiety (Zeidner, 1998).

16. Interference model – this model suggests that test anxiety artificially lowers the performance on cognitive ability tests; measurement bias of test scores exists as a result of test anxiety (Wine, 1971).


18. Self-efficacy – individuals’ conviction in their own competence to attain desired goals in particular domains (Bandura, 1997).
CHAPTER TWO: LITERATURE REVIEW

Introduction

Combing through the literature provided a glimpse into a research history of the test anxiety construct, the evolution of quantifying student motivation, and test anxiety interventions. Test anxiety research has shifted from the belief that the test anxiety construct is unidimensional to the current view that test anxiety is multidimensional. This shift was evidenced by the transformation of test anxiety scales from measuring test anxiety alone as a unidimensional construct, the presence or absence of debilitating test anxiety, to a measure of multiple components of test anxiety, which includes but is not limited to components such as physiological hyperarousal and facilitating test anxiety. A similar evolution in test anxiety intervention strategies took place with evidence pointing to multimodal forms of treatment being more effective than a single, isolated form of treatment (Zeidner, 1998).

Tests are designed with the intent to measure what students know. The increase in high-stakes testing combined with the increase in the prevalence of test anxiety in K-12 students in the United States has also created a sense of urgency not only for identifying which students are most susceptible to the debilitating effects of test anxiety but also for designing effective test anxiety intervention programs. In order to do that, it is vital to examine issues regarding test anxiety and student motivation. The evolution of computerized adaptive tests has not only provided additional data on student achievement but has the added advantage of being able to measure student effort, which evidence has shown that “the effort an examinee devotes to a test may dynamically vary throughout the test” (Wise & Kong, 2005, p. 165).

The increased emphasis on district, school, and teacher accountability as a result of NCLB high-stakes testing is also affected by test anxiety. It can best be summed up by von der
Embse and Hasson (2012), who stated that it was “essential to address any variable which could improve student test scores” (p. 184). Given the prominent nature of high-stakes testing results and the accountability associated with these results, achievement test measures needed to be free from any form of measurement bias (Reeve & Bonaccio, 2008). This was applicable to any test used for accountability whether the test was classified as high-stakes or low-stakes for the test-taker, as the stakes for the test giver may be different than stakes for the test-taker. The objective of this literature review was to synthesize the theoretical frameworks of test anxiety with aggregate response time and response time effort (RTE) in an attempt to explore relationships between those variables.

Early test anxiety research conceptualized the test anxiety construct as being unidimensional (Mandler & Sarason, 1952). This specific concept was not only reflected by the test anxiety research itself but also by the test anxiety scale instruments used to measure test anxiety. A shift in the test anxiety conceptualization occurred in the 1960s when work by Alpert and Haber (1960) and Liebert and Morris (1967) proposed that test anxiety consisted of multiple components and, thus, was multidimensional. Therefore, test anxiety evolved from being viewed as a discrete variable where test anxiety was either present or not present to being a continuous variable whereby the magnitude of its presence was measured (Zeidner, 1998).

The bidirectional theory work of Alpert and Haber (1960) confirmed the existence of a facilitating dimension of test anxiety, a type of anxiety which was shown to increase student achievement scores. Until the work by Alpert and Haber, test anxiety was seen as being unidimensional in that students either did or did not experience debilitating test anxiety. Liebert’s and Morris’ (1967) research, too, focused on multidimensional aspects of test anxiety, specifically worry and emotionality, which were correlated but still considered to be empirically
distinct facets of test anxiety, with worry being deemed as having a more detrimental impact than emotionality on student achievement (Hembree, 1988; Seipp, 1991). It was during the late 1960s and early 1970s that a fundamental shift occurred in test anxiety research such that test anxiety was also viewed as being less of a behavioral construct and more of a cognitive construct (Hembree, 1988). Worry and emotionality have emerged to become the two most prominent dimensions of test anxiety (Hodapp & Benson, 1997).

Worry was the cognitive component of test anxiety whereby students were unable to focus on a task due to task-relevant thoughts and task-irrelevant thoughts and concerns about performance on the task (Wine, 1971). Task-relevant thoughts are thoughts that were directly related to the problems involved in a given task, whereas the task-irrelevant thoughts pertained to the worry students experienced related to themselves and their performance on a given task. Students with low levels of test anxiety were typically task-focused, whereas students with higher levels of test anxiety tended to have task-irrelevant thoughts (Harpell & Andrews, 2013). The presence of worry “decreases the availability of working memory for a cognitive task” (Fritts & Marszalek, 2010, p. 443). Furthermore, this type of anxiety reduced a student’s cognitive ability, which included working memory (Eysenck, Derakshan, Santos, & Calvo, 2007).

Emotionality, which is now referred to as physiological hyperarousal (Joiner et al., 1999), was the emotion-based, or affective component of test anxiety - the physical symptoms, such as perspiration, increased blood pressure, rapid or shallow breathing, and an increased heart rate that occur during testing (Wine, 1971). Hembree’s (1988) meta-analysis found that worry, the cognitive component of test anxiety, was more strongly correlated to students’ test performances than emotionality. The test anxiety relationship to student achievement had a correlation value
of \( r = -.31 \), while the test anxiety’s relationship to emotionality had a correlation value of \( r = -.15 \) (Hembree, 1988). In a subsequent meta-analysis correlation by Ackerman and Heggestad (1997), the correlation value between test anxiety and general intelligence measures was \( r = -.33 \).

Anxiety itself was differentiated between being trait anxiety and being state anxiety. Huberty (2010) defined trait anxiety as “anxiety that is chronic and pervasive across situations and is not triggered by specific events” (p. 13), and state anxiety as being “anxiety that occurs in specific situations and usually has a clear trigger” (p. 13). Trait anxiety was viewed as being a more stable state of anxiety, which was typically associated with anxiety disorders. State anxiety, on the other hand, was seen as being a reflection of the emotionality, or physiological hyperarousal, component of test anxiety (Whitaker Sena et al., 2007). Students with a combination of higher levels of trait anxiety and lower levels of ability were at a greater risk of having higher levels of debilitating test anxiety (Reeve & Bonaccio, 2008).

Measurement of trait anxiety, which was anxiety that is measured separate from a specific assessment situation, and measurement of state anxiety, which was anxiety measured relative to and immediately following a specific assessment, were found to have varied influences on the test anxiety students experienced (Croyle, Weimer, & Eisenman, 2012). It was found that the correlation between test anxiety and academic performance was stronger when test anxiety was measured after an assessment rather than before the assessment (Seipp, 1991). Differentiating between trait anxiety and state anxiety was examined to determine how changing the context of test anxiety measurement affected the aspects of test anxiety being measured. Thus, it was best to examine the test anxiety students experienced either during or following a given assessment (Wolf & Smith, 1995). Interestingly enough, student achievement was found to be a significant predictor, even the strongest association of the variables studied, in the
measurement of state test anxiety but not for the measurement of trait anxiety, which lends support for a “self-regulative model of test anxiety” (Croyle et al., 2012, p. 11). Thus, the findings of test anxiety studies should include the context of when test anxiety was measured to alleviate potentially misleading interpretations of the resulting data.

Anxiety may be manifested in a variety of ways, such as behaviorally, cognitively, and physiologically, which can be evident by a variety of primary characteristics (Huberty, 2010). Furthermore, Wren and Benson (2004) posited that test anxiety was comprised of three components, which are Thoughts, Automatic Reactions, and Off-Task Behaviors. Recent research by Lowe et al. (2008) proposed additional dimensions of test anxiety, which included four debilitating test anxiety subscales. The debilitating scales were Cognitive Obstruction/Inattention, Physiological Hyperarousal, Social Humiliation, and Worry. In addition, there was Performance Enhancement/Facilitation Anxiety, which was a facilitating test anxiety scale, and a Lie scale, which helped identify the truthfulness of students’ responses on a self-report test anxiety survey (Lowe et al., 2008).

The symptoms experienced by students with higher levels of test anxiety typically varied by age with younger students displaying more physical symptoms and older students displaying more cognitive symptoms (Whitaker Sena et al., 2007). Since then more research was conducted with elementary and middle school students, which led some researchers to hypothesize that facilitating test anxiety may vary by age (Lowe & Lee, 2008). This was in line with previous research that indicated that the debilitating effects of anxiety tended to increase through the middle school grade levels, reached a peak in the early high school grade levels, and leveled off for high school juniors and seniors (Hembree, 1990). It was noted, however, that younger
students instead were more susceptible than older students to higher levels of worry as a result of academic stress students experienced in school settings (Harpell & Andrews, 2013).

The combination of the worry and physiological components of test anxiety has been shown to have an inverse relationship with student achievement. The resulting increase in test anxiety was also positively correlated to cognitive distortions in both the social and academic domains (Putwain et al., 2010a). A potential confounding variable of test anxiety was the anxiety associated with a particular content area, such as mathematics, where a negative relationship between math anxiety and math achievement was also found (Ma, 1999). Hembree (1990), however, purported that “a tacit belief has seemed to prevail that test anxiety theory can be used to support both (test anxiety and math anxiety) constructs” (p. 34). Regardless of how anxiety is manifested, a negative association between anxiety and academic achievement has been well established (Ergene, 2003; Hembree, 1988; Owens, Stevenson, Norgate, & Hadwin, 2008; Peleg, 2009; Rezazadeh & Tavakoli, 2009; von der Embse, Barterian, & Segool, 2013).

**Theoretical Frameworks**

Various theoretical frameworks were utilized for test anxiety research. Social learning theory and expectancy theory were among the frameworks cited in test anxiety research, as well as a variety of test anxiety models, which included the following: deficit, interference, and transactional. Due to the multidimensional and, thus, heterogeneous nature of test anxiety, however, its complexity could not be adequately described by one particular theoretical perspective (Zeidner, 1998).

**Social Learning Theory**

The model of learning espoused by Bandura (1986) included three interdependent components, otherwise known as the triadic reciprocal causation: a) a person’s biological and
psychological characteristics, b) a person’s behavior, and c) the person’s environment. Aspects of social learning theory, such as self-efficacy, which was the conviction one has in his or her ability to be successful, were used to explain differences in student performance in conjunction with test anxiety. Academic and social behaviors were impacted by self-efficacy, which “affects all types of behavior – academic, social, and recreational” (Miller, 2011, p. 244). Three environments impacted one’s self-efficacy. These environments included the following: a) imposed, b) selected, and c) created (Bandura, 1997).

In a study by Collins (1982), students with high or low levels of mathematics self-efficacy were given challenging math problems to solve. Students with higher levels of self-efficacy displayed higher levels of persistence, which implied their attribution of performance to effort, whereas the students with lower levels of self-efficacy felt that their performance was a reflection of their low ability in mathematics. Bandura (1986) encouraged providing students with high levels of test anxiety with frequent successful testing experiences in an effort to increase students’ self-efficacy levels for subsequent tests.

In a study of the multidimensionality of test anxiety, it was stated that “self-efficacy cannot be subsumed under the concept of test anxiety” (Hodapp & Benson, 1997, p. 239). In addition, research results were mixed regarding the association between self-efficacy and student achievement. Nielson and Moore (2003) touted that student achievement scores were related to students’ academic self-efficacy. Recent research showed that a significant correlation existed between students’ self-efficacy beliefs and student achievement (Abdi, Bageri, Shoghi, Goodarzi, & Hosseinzadeh, 2012). Keith, Hodapp, Schermelleh-Engel, and Moosbrugger (2004) contended that test anxiety was separate from self-efficacy and was instead related to self-esteem, which previously was found to have a strong inverse relationship with test anxiety.
(Hembree, 1988). Self-esteem was defined as being a combination of self-efficacy and a lack of confidence, a construct which was found to be highly correlated to test anxiety (Hodapp & Benson, 1997). Self-esteem was also said to have a positive correlation to academic achievement (Peleg, 2009).

A meta-analysis by Preiss, Gayle and Allen (as cited in Putwain, 2008a, p. 147), which examined test anxiety, self-efficacy, and study skills, noted the existence of an inverse relationship between self-efficacy and test anxiety such that students with higher levels of test anxiety had lower levels of self-efficacy. Inverse relationships were found to exist not only between test anxiety and self-esteem but also between test anxiety and self-efficacy (Pekrun et al., 2004). Self-esteem and self-efficacy were determined to predict test anxiety measured in both trait anxiety and state anxiety contexts (Croyle et al., 2012). There was, however, disagreement with these findings, as it was also suggested that no significant correlation existed between self-efficacy and test anxiety (Abdi et al., 2012). Students with higher levels of test anxiety “usually have low levels of self-efficacy” (Abdi et al., 2012, p. 419), but students’ self-efficacy was not deemed to be an effective predictor of students’ test anxiety. Although students with high test anxiety levels tended to have low self-efficacy levels, that did not mean that students with low levels of test anxiety had high levels of self-efficacy, which would then indicate a correlation. It was suggested that there was a complex relationship between test anxiety and self-efficacy and, thus, due to this complexity the relationship may potentially be mediated by other variables (Chamorro-Premuzic, Ahmetoglu, & Furnham, 2008).

Galla and Wood (2012) found that a student's level of emotional self-efficacy, or one’s ability to regulate negative emotions, had a moderating effect on test anxiety. The math achievement scores of students with high levels of test anxiety and with high levels of emotional
self-efficacy were less affected than the math achievement scores of students with high levels of test anxiety but low levels of emotional self-efficacy. Thus, the implication was that the anxiety experienced only by the students with low levels of emotional self-efficacy was negatively correlated with their math achievement scores. This would suggest that self-efficacy was a controlled for variable. In addition, higher levels of emotional self-efficacy may serve as a mediator of the negative impact anxiety has on math achievement scores (Galla & Wood, 2012).

Previous research has shown that math self-efficacy may reduce test anxiety and perhaps increase student achievement (Bandalos & Yates, 1995). Self-efficacy has been found to have a significant correlation with academic achievement but not necessarily with test anxiety, which suggests that self-efficacy is not a generally reliable, effective predictor of test anxiety; however, self-efficacy, academic achievement, and test anxiety may not necessarily be mutually exclusive (Abdi et al., 2012). It was shown that the level of a student’s certainty of his or her achievement performance has an influence on the relationship between the level of expectancy of success or failure and the degree of the student’s preparation and persistence (Dickhäuser, Reinhard, & Englert, 2011). The student achievement-test anxiety relation, the student achievement-worry relation, and the student achievement-emotionality relation have each been found to be mediated by academic self-concept as well, which can be improved by providing students with individual feedback, rather than social or peer group feedback frames of reference (Goetz, Preckel, Zeidner, & Schleyer, 2008).

While anxiety does have a negative effect on student achievement, the magnitude of the effect that anxiety has on each student individually does in fact vary based on each student’s self-efficacy. Bonaccio and Reeve (2010) noted that self-efficacy also varies by age, with younger students being more susceptible to low self-efficacy than older students. Academic self-concept
was also found to vary by age such that older students tended to have higher levels of academic self-concept (Harpell & Andrew, 2013). Furthermore, the lack of students’ self-efficacy has been shown to be a contributing factor, albeit the relatively least important factor, of the sources of test anxiety as perceived by students (Bonaccio & Reeve, 2010). Given the pervasive emphasis on student achievement, teachers are typically more skilled at determining student achievement performance than judging affective areas, such as student motivation and test anxiety levels (Urhahne et al., 2011). Hence, this would suggest a need for test anxiety measurement resources to help teachers better identify students who have experienced high levels of debilitating test anxiety.

The interaction effects of valuing success and a student’s expectation for success support self-esteem’s place in the test anxiety realm (Selkirk, Bouchey, & Eccles, 2011). Improving the academic self-concept of students was a suggested intervention to reduce the school stress and reduce cognitive interference students experience (Harpell & Andrews, 2013). Improving students’ academic self-concept is a noteworthy strategy, as it was found that academic self-concept served as a mediator of the relationship between test anxiety and student achievement (Goetz et al., 2008). It was shown that by improving students’ “self-perceived competence” (Lang & Lang, 2010, p. 812), students who previously suffered from high test anxiety levels experienced both lower test anxiety levels and an increase in student achievement. Thus, the perceptions students have of themselves affected both academic performance and test anxiety levels (Urhahne et al., 2011).

Spinath noted that while teachers were moderately adept at estimating a student’s academic self-concept, teachers’ accuracy for estimating students’ motivation to learn and accuracy for estimating students’ test anxiety levels were much lower (as cited in Urhahne et al.,
It is also possible that students’ academic self-concept may be influenced by the kind of class students are in. For example, gifted students in gifted classes were found to have lower academic self-concept levels and elevated levels of test anxiety in comparison to gifted students in regular education classes (Goetz et al., 2008).

It is worth noting that, due to findings in his meta-analysis, Hattie (2009) touted the use of interventions to provide additional support to students whom teachers underestimated in terms of academic achievement. Underestimated students tended to have lower levels of academic self-concept, lower levels of motivation, and higher levels of debilitating test anxiety. Underestimated students also lacked self-efficacy (Bandura, 1997). In addition, lower levels of academic self-concepts were shown to be significant predictor variables of both cognitive interference and a lack of confidence for students (Harpell & Andrews, 2013). Lack of confidence was found to be a salient dimension of test anxiety (Hodapp & Benson, 1997). Although it has been established that a significant positive correlation exists between academic self-concept and student achievement, academic self-concept may not necessarily be a reliable predictor of student success in school (Hattie, 2009). Nonetheless, students’ self-esteem cannot be discounted; students bring to their testing experience not just their academic self-concept but also how they feel about themselves.

**Expectancy Theory**

Vroom (1964) touted that a student’s expectancy of achieving a performance outcome influenced the effort he or she put forth in order to reach that outcome. It has typically been implied that test scores on large-scale assessments represent students’ best efforts; yet without an assurance that the test itself is valued by the student, it is difficult to know to what extent test scores were influenced by students’ efforts and, subsequently, their motivation. According to
Wainer (as cited in Wolf & Smith, 1995), if there were no consequences associated with test-takers’ test scores, it was possible that the test scores were influenced by students’ lack of effort; and therefore, the observed test scores were not truly representative of what students know.

An interaction effect exists between expectancy and the level of certainty a student has about meeting or failing to meet that expectancy. This interaction effect enjoins social learning theory to expectancy theory. When a student was confident and expected to do well, his or her expectancy influenced both the level of preparation prior to the test and the degree of perseverance during the test. Students who were highly certain to meet goals remained steadfastly persistent, while students who believed they were certain to not meet goals were not persistent, which, subsequently, impacted student achievement (Dickhäuser et al., 2011). The results from the Dickhäuser et al. (2011) study further suggested that certainty should be included as an “additional variable in expectancy models” (p. 519). Wise and DeMars (2005) found that the relationship between student effort and student achievement on assessments had an average effect size which exceeded 0.5 standard deviation. This finding was comparable to the meta-analysis by Seipp (1991), which indicated a difference of almost half of a standard deviation between the achievement of students with low levels of test anxiety and those students with high levels of test anxiety.

Furthermore, the difference in student achievement between the lowest response time effort (RTE) group and the highest RTE group was found not only to be significant, but it also exceeded two standard deviations (Wise & Kong, 2005). In the expectancy-value model of achievement motivation, Wise and DeMars (2005) related student achievement to students’ perceived expectations of success as well as the students’ perceived value of doing well on a
given assessment. Student effort can then be examined at two levels: a) the aggregate level (total duration of the assessment), as well as b) the individual item level (Setzer et al., 2013).

The effort of students on a test was also deemed to potentially be a “serious potential threat to test score validity” (Wise & Kong, 2005, p. 163). Students may not give their best effort if there are no consequences for having poor test scores or there are no benefits for students with high test scores. The extent of effort may be based on factors such as the quality of students’ work and the amount of time students spend taking an assessment. This potentially places a greater burden on educators to ensure that students are engaged, motivated, and putting forth their best effort on assessments. Furthermore, it was demonstrated that the components of effort and task-values were “significant predictors of ... academic achievement in mathematics” (Zerpa, Hachey, van Barneveld, & Simon, 2011, p. 7).

Ryan, Ryan, Arbuthnot, and Samuels (2007) touted that NCLB assessments were predicated on a “unidimensional view of motivation” (p. 11). Past research by Budescu and Bar-Hillel (1993) also differentiated students by either being ideal test takers, which are students who have a goal to “maximize test performance” (as cited in Ryan et al., 2007, p. 11), or real test takers, which are students who may have goals other than maximizing their performance. Previous findings indicated that the best performances on tests were by students with high levels of motivation and low levels of test anxiety while the worst performances on tests were by students with low levels of motivation and high levels of test anxiety (Wolf & Smith, 1995). The difference in motivation may be potentially affected by the instrumentality of the test itself. The instrumentality, in test anxiety terms, is what the student perceives the purpose of a test to be. Zohar and Brandt (2002) posited that a test’s instrumentality, or purpose, positively correlated with the amount of test anxiety a student experienced.
Reeve, Bonaccio, and Charles (2008) noted that the instrumentality, or purpose, of a test, when differentiated as being primarily evaluative or developmental, had “the greatest effect on a person’s expected test anxiety” (p. 246). This finding was also confirmed by the heightened test anxiety levels students experienced on NCLB assessments when compared to the test anxiety levels students experienced when taking classroom assessments (Segool et al., 2013). When students were classified as having low, medium, or high test anxiety levels, those classifications were able to be predicted based on the students’ achievement on a high-stakes assessment, which suggested that test anxiety does impact performance (von der Embse et al., 2014). This indicated a need for more research to establish criteria to classify individuals according to their test anxiety levels.

Furthermore, the instrumentality of a test and a student’s self-confidence, or how confident one is of doing well on a test, were shown to be more important factors than familiarity with the testing format and what the test is actually measuring, such as a skill or a student’s ability, in determining a student’s expected anxiety (Reeve, Bonaccio, & Charles, 2008). Self-confidence was found to have the second most significant effect on test anxiety in the study. It was also suggested that self-confidence be incorporated as an additional factor as part of multidimensional test anxiety inventories (Peleg, 2009). In a subsequent study by Bonaccio and Reeve (2010), the instrumentality, or the purpose of the test, was the second highest perceived source of test anxiety, while the highest perceived source of test anxiety was attributed to students’ perception of the difficulty of the test itself. As evident in the aforementioned studies, the instrumentality, or purpose, of a test does play a significant role in the test anxiety levels of students.
A factor in the expectancy-value model of achievement motivation was the effort students expended while taking a test, effort that is oftentimes dependent on the stakes of the assessment itself (Wise & DeMars, 2005). The expectancy-value model was applicable to both a test as well as individual test items. The expectancy-value interaction, whereby a student’s performance is a product of both the performance a student expects of himself or herself and what he or she values, was also a salient factor in the test anxiety construct. Students who placed a high value on success but have low performance expectations in English or math tended to have the highest levels of test anxiety, with the effect size being larger in math than in English (Selkirk et al., 2011). A caveat to consider in light of the expectancy-value interaction was that while increasing students’ expectancy levels of success is a worthy goal, ensuring that students also do not overvalue a particular subject area, such as English or math, was also significant, since overvaluing could potentially lead to an increase in test anxiety and may subsequently have a negative effect on student achievement, particularly for low achieving students (Selkirk et al., 2011).

Although improving a student’s self-confidence may reduce test anxiety and potentially improve performance, a delicate balance is needed between improving the self-confidence of students’ with high levels of test anxiety and increasing the self-confidence of students with low levels of test anxiety. It was found that increasing the self-confidence of a student with low test anxiety may actually have a detrimental effect on performance due to a student becoming overconfident, which may lead to a decline in focus and effort and, subsequently, a poorer achievement score, which is in line with expectations based on disjunctive motivational models (Lang & Lang, 2010).
Related Literature

Models of Test Anxiety

Each model that follows addresses various aspects of test anxiety that impact student achievement. Many test anxiety theoretical perspectives generally contain facets of either the deficit model or the interference model (Bonaccio & Reeve, 2010).

Deficit model. The deficit model proposes that test anxiety does not directly contribute to lower student achievement scores but rather that individuals who have lower ability levels tended to suffer from higher levels of test anxiety, which supported the correlation between test anxiety and student achievement scores (Covington & Omelich, 1987). Zeidner (1998) posited that the correlation was not causal whatsoever but was simply an “epiphenomenon reflecting students’ lack of preparation for the test and their meta-cognitive awareness of their low probability of succeeding on the exam” (p. 70). Hembree’s (1988) meta-analysis purported that an inverse relationship exists between students’ test anxiety levels and their intelligence quotient (IQ) such that students with higher levels of test anxiety tended to have lower IQs, although this relationship could potentially be confounded by test anxiety. In the deficit model, students with high levels of test anxiety and low student achievement scores attributed their academic performance to poor study habits or poor test-taking strategies rather than being a result of test anxiety.

Zeidner (1998) posited that several deficit models existed. For example, one such deficit model was that students with lower ability levels simply report higher levels of test anxiety, while a second deficit model was that poor study skills resulted in poor performance and, thus, increased test anxiety. Therefore, deficit models implied that test anxiety did not influence achievement scores and, thus, no measurement bias existed as a result of test anxiety;
achievement scores were believed to be an accurate measurement of student performance, regardless of the test anxiety experienced by testtakers (Zeidner, 1998). The Yerkes-Dodson Law, which touted a curvilinear relationship between anxiety levels and performance, may be represented in a way such that the test anxiety experienced by students interacted with their ability, which would support a deficit model framework (Wicherts, Dolan, & Hessen, 2005).

Reeve and Bonaccio (2008) posited that students with lower ability levels experienced higher levels of test anxiety, which implied that tests “measure ability equivalently across levels of anxiety” (p. 534). Thus, the relationship between test anxiety and student achievement was due to differences in ability rather than measurement bias. Furthermore, it was even suggested that efforts to reduce test anxiety through various means of test anxiety interventions may actually induce measurement bias on assessments (Reeve & Bonaccio, 2008). Such a stance was in stark contrast to Hembree’s (1988) meta-analysis, which supported the interference model.

When test anxiety was shown to have a negative impact on criterion-related validity in Classical Test Theory (CTT), it was also found to be highly variable (Reeve, Heggestad, & Lievens, 2009). Not all research demonstrated that test anxiety has a negative impact on test validity, and, in fact, the effects of test anxiety regarding cognitive ability test performance may actually enhance test validity, as demonstrated by a .21 correlation between test scores and test anxiety (Wicherts & Scholten, 2010). This, however, was only shown to be true in low-stakes assessment settings and could not be generalized to high-stakes assessment settings.

**Interference model.** Wine (1971) contended that students experienced a form of cognitive interference, or obstruction, which, in turn, resulted in artificially lowered student achievement due to test anxiety’s impact on measurement bias of test scores. The cognitive interference that students with high levels of test anxiety may have experienced was shown to
interfere with student achievement based on the difficulty level of a single test item such that recall for an easy item was interfered with more significantly than recall for a difficult test item (Covington & Omelich, 1987). An inverse relationship between test anxiety and student achievement was noted as part of a meta-analysis such that higher test anxiety correlated to lower student achievement (Hembree, 1988), while more recent findings have suggested that the relationship between test anxiety and student achievement may be nonlinear, which would be a reflection of the Yerkes-Dodson Law, which stated that test anxiety, debilitating or facilitating, was moderated by the difficulty of a test (Szafranski, Barrera, & Norton, 2012). Seipp (1991) noted in her meta-analysis that “two-thirds of the low-anxious students score better than the average high-anxious student” (p. 38). The essence of the interference model is that higher levels of test anxiety impact student achievement by “competing for cognitive resources” (Bonaccio, Reeve, & Winford, 2012, p. 498).

The interference model consists of the task-irrelevant thoughts students experience when taking a test. Steele and Aronson (2000) touted Stereotype Threat as possibly the most significant example of an interference perspective. Stereotype Threat (ST) is the perceived threat that a particular subgroup may experience which creates interference that results in student achievement levels that are lower than what represents a student’s actual aptitude. In turn, this perceived threat leads to a form of measurement bias, as results are not representative of a student’s actual knowledge (Wicherts et al., 2005).

Students who identify with groups that may be marginalized as determined by factors, such as gender, ethnicity, special education, and low socioeconomic status, may be negatively impacted by ST (Fritts & Marszalek, 2010). Students from culturally and linguistically diverse backgrounds have typically had higher levels of debilitating test anxiety, in part, due to ST
(Osborne, Tillman, & Holland, 2010). For example, it was found that merely having African-American students identify their race on a test could lead to students in that particular group to answer fewer questions and with more incorrect responses than their Caucasian peers (Steele & Aronson, 2000).

The cognitive interference students experienced was shown to mediate the relationship between student achievement and the two facets of test anxiety, worry and emotionality, which also found that the correlation between test anxiety was stronger with the worry component of test anxiety than with the emotionality component of test anxiety for all students (Putwain et al., 2010a) and even student subgroups, such as students with disabilities and females (Datta, 2013). A negative correlation was found between test anxiety and cognitive function such that students with lower intelligence quotients had high levels of test anxiety (Hembree, 1988). Whitaker Sena et al. (2007) also found that for students with disabilities, cognitive interference was the strongest predictor of test anxiety.

Predictive validity of cognitive ability assessments was shown to decrease as worry increased; however, the predictive ability of assessments seemed to be unaffected by emotionality, which would indicate the presence of cognitive interference (Bonaccio et al., 2012). The authors of this research further cautioned practitioners in the appropriate use of resulting data of cognitive ability assessments such that prediction of performance on such tests should only be used for students with low test anxiety levels. The differential validity indicated that the prediction accuracy for students with high levels of test anxiety was less valid as the data for students with high levels of test anxiety may contain more decision errors and may not truly reflect those particular students’ cognitive skills (Bonaccio et al., 2012).
Putwain (2007a) poignantly stated that “if (student achievement) performance was not affected (by test anxiety), then it is unlikely that test anxiety would have ever become a phenomenon of educational … study” (p. 151). Recent research indicated that a significant correlation exists between cognitive functioning and test anxiety (Abdi et al., 2012). Furthermore, it was even suggested that worry interference and cognitive interference should be studied as separate distinct facets of the test anxiety cognitive domain (Harpell & Andrews, 2013). Having a deeper understanding of how test anxiety affects students cognitively will better inform educators about which interventions (von der Embse et al., 2013) to use and how to more effectively target the nature of the cognitive interference students are experiencing.

**Transactional model.** In the transactional model, anxiety occurred as a result of the perceived stress one experiences when one felt that his or her coping ability was insufficient to adequately cope with a stressful situation (Lazarus & Folkman, 1984). The focus of the transactional model was on the interaction between the environment and students’ personality traits. Within a transactional model, the test anxiety a student experienced was a result of the student determining that his or her coping resources were inadequate to handle the test anxiety he or she was experiencing (Harpell & Andrews, 2013). This is an example of state anxiety, which is dependent upon how an individual responds to perceived threats. This is in contrast to trait anxiety, which is a more stable proneness to anxiety (Spielberger & Vagg, 1995). Students who exhibited high levels of test anxiety were typically prone to suffer from elevated levels of trait anxiety, which manifested itself when students viewed an evaluative situation as threatening and, thus, led to varying degrees of state anxiety (Lowe et al., 2008).

One focus of the transactional model in regard to test anxiety was on the students’ perceptions of the testing situations they experienced or the perceptions students had of
themselves in relationship to the testing situations they experienced. Perceived sources of test anxiety may include, but are not limited to, things such as how the test scores will be used as a factor in making academic decisions for the student, the purpose or instrumentality of the test, and students’ perceptions of the complexity or difficulty of the test. Test difficulty and instrumentality of the test endorsed by students’ perceptions were perceived as being the first and second most important sources of test anxiety, respectively (Bonaccio & Reeve, 2010). The transactional model places an emphasis on the learning and testing environment as well as the social supports essential to addressing test anxiety with an effort to increase students’ confidence levels and, subsequently, boost student achievement scores (Harpell & Andrew, 2013).

**Test Anxiety Models Summary**

Test anxiety researchers have not arrived at a consensus regarding which model is most salient, as test anxiety is viewed as being a complex, multidimensional construct. The evidence in Hembree’s (1988) meta-analysis supported the interference model over the deficit model. The three implications which resulted from Hembree’s (1988) work are as follows: (a) the performance of students suffering from high levels of test anxiety is oftentimes misinterpreted and undervalued, (b) student achievement scores are inherently biased at the national, state, and local levels, and (c) the “validity of the entire testing process is challenged” (p. 75). Thus, measurement bias was a factor not only in student achievement but also in the very quality of assessments themselves. When controlling for worry, the cognitive component of test anxiety, it was found that the correlation between emotionality and test performance became negligible (Cassady & Johnson, 2002; Hembree, 1988).

Wichert and Scholten (2010) evaluated Stereotype Threat (ST) within a measurement framework, which revealed that ST effects were a source of measurement bias, a result which
was consistent with the interference model. However, this was refuted by Reeve and Bonaccio (2008), who indicated that student achievement was not directly influenced by test anxiety. This finding suggested that lower ability was associated with higher levels of test anxiety, while higher ability was associated with lower levels of test anxiety, a finding that was consistent with the deficit model. Furthermore, when test anxiety was induced via stereotype threat in a low-stakes setting, the convergent validity of subtests among various treatment groups was the highest when factoring in student ability (Wicherts & Scholten, 2010).

While the interference model and deficit model each purport distinctly different relationships between test anxiety and the manner of its impact on student performance, what the aforementioned models do agree on is the existence of test anxiety. Although a vast majority of test anxiety research focused on the interference model and the deficit model, the transactional model, with an emphasis on environments, has a research following and has a role in the study of test anxiety. As Zeidner (1998) noted, all test anxiety models have made at least some reference to test anxiety having an effect on performance. What is in question is how and to what extent that test anxiety impacts student achievement.

**Effects of Test Anxiety on Student Groups**

Test anxiety has been shown to affect various student groups in different ways. Females tend to exhibit higher levels of debilitating test anxiety than males, and students with disabilities exhibit higher levels of debilitating test anxiety than their peers without disabilities. Test anxiety has also been shown to vary by ethnicity and also by socioeconomics. The test anxiety of the entire population sample, as well as the test anxiety by gender, was examined in this study.

**Gender.** Test anxiety has different effects on different student populations. Studies have shown that females tended to experience more test anxiety than males (Cassady & Johnson,
2002; Datta, 2013; Fritts & Marszalek, 2010; Harpell & Andrews, 2013; Hembree, 1988; Lowe, Grumbein, & Raad, 2011; Lowe & Lee, 2008; Segool et al., 2013). Hembree’s (1988) meta-analysis further noted that there was a significant difference in debilitating test anxiety between males and females at all grade levels, including post-secondary environments, such that females experienced higher levels of debilitating test anxiety than males. Renorming the Test Anxiety Inventory (TAI), which was first normed in 1980 with undergraduate students, indicated that test anxiety scores among females had increased in the past 30 years while test anxiety scores among males had held constant (Szafranski et al., 2012).

While females typically experience greater levels of test anxiety, particularly worry, than males, males were shown to have higher levels of “behavioral manifestations of School Stress” (Harpell & Andrews, 2013, p. 80). In addition, the physiological symptoms exhibited by males are more visible than the worry symptoms that females experience, which may lead to educators focusing more on addressing the test anxiety experienced by males than the test anxiety experienced by females. Stereotype threat was identified as a potential mediating variable and may potentially explain the phenomena of females experiencing higher levels of test anxiety than males (Fritts & Marszalek, 2010).

Given the evolution of test anxiety from unidimensional to bidimensional and, subsequently, multidimensional, it is worth noting that test anxiety differences by gender have been observed among each of the various multidimensions as measured by more recent test anxiety scales with females experiencing elevated levels of debilitating test anxiety when compared to males, while males had slightly higher scores on the lie scale, a validity construct of participants’ true or ideal test behavior (Lowe & Lee, 2008). However, gender differences have not always been evident in each individual factor of test anxiety, such as no significant difference
between genders being evident in the social concern/humiliation factor dimension (Lowe et al., 2011). Furthermore, no significant difference between genders was observed in the performance enhancement/facilitation test anxiety factor, the factor of test anxiety which enhances student achievement (Lowe & Lee, 2008). This result was contrary to what Hembree’s (1988) found in meta-analysis which indicated that statistically significant differences in facilitating test anxiety existed between males and females with males having higher levels of facilitating test anxiety than females. It is noteworthy to mention, though, that most of the studies in Hembree’s meta-analysis consisted of students in grades 9-12 and students in college.

Although females reported higher levels of test anxiety than males when taking assessments in the classroom or online, no statistically significant difference was found when controlling for the setting of the assessment, whether the assessment was taken online or taken in the classroom (Stowell & Bennett, 2010). Although the effect size of the difference was small, females experienced higher levels of debilitating test anxiety than males not only when taking classroom assessments but when taking NCLB assessments as well (Segool et al., 2013).

**Students with disabilities.** Many studies have indicated that students with disabilities are more prone to experiencing higher levels of test anxiety when compared to students who do not receive special education services (Whitaker Sena et al., 2007). Students with disabilities were susceptible to not only higher levels of test anxiety but also higher rates of test anxiety prevalence than their peers without disabilities (Datta, 2013; Salend, 2011a). Furthermore, it was also noted that students with disabilities had lower levels of self-esteem than their peers without disabilities (Peleg, 2009). Thus, accessibility to test content is paramount for all students. For students with disabilities it is necessary to ensure that students are provided appropriate and possibly a wide range of accommodations in the classroom and on classroom
tests throughout the school year as well as assigning the proper NCLB assessments to each student with special needs. Salend (2011b) proposed placing greater emphasis on the use of valid tests that are student-friendly by making tests more accessible to students by “improving directions, format, readability, and legibility” (p. 54).

Students with learning disabilities are more likely to have lower facilitating test anxiety levels as well. Students with learning disabilities had higher worry scores, which was more detrimental to student achievement than emotionality given that worry impacts cognitive processing while emotionality impacts physical symptoms (Whitaker Sena et al., 2007). Not all studies have confirmed these findings, however, as students with disabilities were not found to have higher levels of test anxiety when compared to their regular education peers when taking classroom assessments or NCLB-mandated state assessments (Segool et al., 2013). The kind of disability may also be a factor in emotionality and worry scores, as it was found that students with vision impairments actually had higher emotionality scores rather than worry scores, which was in contrast to other test anxiety research findings (Rezazadeh & Tavakoli, 2009).

**Ethnicity.** It has been shown that African-American and Hispanic students experience more test anxiety than Caucasian students (Hembree, 1988). While much of test anxiety research indicates higher debilitating test anxiety levels for minority students when compared to their Caucasian peers, results have been mixed with some studies showing no statistical difference in test anxiety between student groups based on ethnicity (Yager, 2008). In addition, students’ test anxiety levels did not differ by ethnicity when they had taken classroom assessments compared to when they had taken NCLB assessments (Segool et al., 2013). Thus, it is quite possible that the difference in test anxiety may be due to individual differences rather than a group difference.
**Socioeconomics.** Students with lower socioeconomic backgrounds were found to exhibit higher levels of test anxiety than students from higher socioeconomic backgrounds (Hembree, 1988). A study by Putwain (2007) that identified differences in trait test anxiety supported the same finding in testing students based on their socioeconomic status. Significant relationships between test anxiety subscales and student achievement were found based on students’ socioeconomic status (von der Embse & Hasson, 2012). Many test anxiety studies have indicated that future research studies of test anxiety should more closely examine differences based on socioeconomic status. For example, on a high-stakes math assessment in Ohio, socioeconomics accounted for 53% of the student performance variance, while test anxiety accounted for 15% of the variance (von der Embse & Hasson, 2012).

**Online Assessments and Paper-Based Assessments**

Some believe that the increased use of computers as a means to assess students presents a new testing format that may potentially affect test anxiety levels for students. While some concerns have been raised about the effect online testing may have on test anxiety when compared to a paper and pencil testing format, Fritts and Marszalek (2010) determined that computer use anxiety differences were minimal and not a factor in student performance regardless of the testing format being used. While in some instances test anxiety levels were found to vary when comparing online testing formats to paper and pencil formats, controlling for demographic factors, such as socioeconomic status and ethnicity, had not always been taken into consideration and should be further examined (Fritts & Marszalek, 2010).

On the contrary, the comparability of computer-based and paper-administered tests was found to be a potential factor in student performance on high-stakes assessments. While the grade level was not found to have an effect on comparability, the content area being assessed
may be affected such that computer-based assessments were deemed to be more advantageous than paper-based assessments in Language Arts and social studies, while paper-based assessments provided a slight advantage for mathematics assessments (Kingston, 2009). Furthermore, computer-based assessments improved quality assurance, improve test validity, and allowed for timelier reporting of student achievement scores (Kingston, 2009).

While no statistically significant differences were found between classroom paper and pencil quiz scores and online quiz scores, the relationship between test anxiety and test performance was weaker for online assessment than for classroom paper and pencil tests (Stowell et al., 2012). Furthermore, students with high test anxiety in the classroom had lower test anxiety levels and higher test scores when taking assessments online, and students with low test anxiety in the classroom had higher levels of test anxiety and lower test scores when taking assessments online. This implied that the setting in which students took an assessment may have affected not only test anxiety levels but also student achievement on the given assessment (Stowell & Bennett, 2010). This would lend support to the interference model of test anxiety.

In light of these findings, one consideration could be to provide students who have experienced high levels of debilitating test anxiety in the classroom the opportunity to take assessments online and for students who experienced high levels of debilitating test anxiety when taking online assessments to have the opportunity to take assessments in the classroom. An intriguing interpretation of this result was that the range of students’ test anxiety levels was reduced when students took online assessments compared to their test anxiety levels when taking classroom assessments (Stowell & Bennett, 2010). This did not take into consideration any potential differences in test scores that could be attributed to the online versus paper and pencil format. In most cases, however, there were no statistically significant differences in test scores
when student performance on a paper and pencil assessment was compared to student
performance on an online assessment (Paek, 2005).

Nonetheless, as more and more states move to computer-based high-stakes assessments,
such as the Common Core assessments administered by the Partnership for Assessment of
Readiness for College and Careers (PARCC) and the Smarter Balanced Assessment Consortium
(SBAC), students should be given multiple opportunities to become familiar with the high-stakes
online testing systems before the start of operational testing. This would also provide school
districts with opportunities to ensure that local networks are adequate, as network problems can
negatively impact student performance (Kingston, 2009).

One distinct advantage for timed computer-based tests is the ability to identify solution
behavior, which is when students do their best to correctly answer an item, and rapid-guessing
behavior, which occurs when students answer an item quickly (Wise & Kong, 2005). Students’
motivation can then be differentiated between good effort and poor effort. Answers given during
rapid-guessing behavior tended to be random and the correctness of answers were near chance
levels (Wise & Kong, 2005). Although rapid-guessing behavior is typically associated with
speeded assessments and believed to be more prevalent during the latter portion of a speeded
assessment, Wise and Kong (2005) also found that rapid-guessing behavior occurred throughout
testing sessions and that the amount of time students spent on a given item was strongly related
to the amount of reading required for each particular assessment item.

On a low-stakes assessment taken by university students, faster test-taking speed on both
individual items and the overall test was found to be related to lower test scores with correlations
of $r = .71$ and $r = .716$, respectively. This finding contradicted the widely held belief that highly
proficient students had both higher test scores and faster response times, which would have been
indicated by lower correlation values (Silm et al., 2013). The difference in the amount of time students spent who incorrectly answered an assessment item and the amount of time students spent who correctly answered the same assessment item was statistically significant with the greatest differences being found in math and spatial reasoning. The performance on the assessment was found to be influenced by two test-taking effort factors: a) the number of items attempted by students, and b) the mean time students devoted to each item on the assessment. Furthermore, the relationship between students’ response time effort (RTE) and test scores was very close to being linear in nature (Silm et al., 2013).

Further research must be done with online assessments, including controlling for the timing of when the online assessments are administered to students, as online assessments for courses are oftentimes administered on a flexible schedule rather than the predetermined times that classroom assessments are taken by students (Stowell & Bennett, 2010). Given the moderate correlation between classroom test anxiety and online test anxiety students experience, a cautious approach should be taken regarding the generalization of the test anxiety students experience in the classroom to the test anxiety students experience in other settings, such as online assessments.

**Measures of Academic Progress® (MAP®)**

One of the more commonly known computerized adaptive tests (CAT) is the Measures of Academic Progress® (MAP®) tests, which are unspeeded tests developed by the Northwest Evaluation Association™ (NWEA™). While there are a variety of MAP® tests offered, math and reading are generally 50 items and 40 items in length, respectively. Students’ scores are expressed as scale scores called RIT (Rasch Unit) scores that assess student growth at various
times throughout the year as well as across grade levels. The standard errors of scores for math and reading are 3.0 and 3.2, respectively (Wise et al., 2010).

The algorithm used for computerized adaptive tests (CATs), like the Measures of Academic Progress® (MAP®), was designed to assign test-takers items in such a way that a student will have a 50% chance of getting a test item correct when the student exhibits solution behavior, behavior deemed as being effortful on each item based on a predetermined response time threshold. When the student exhibits rapid-guessing behavior, he or she will be correct with a percentage rate that is akin to random guessing (Wise et al., 2013). This does presume, however, that the answer choices for a given assessment are standardized and that the locations of the correct answer choices are balanced across a set of questions. If a middle bias exists in a test where a higher number of correct answer choices are answer choices located in the middle of the multiple choice options, then it would be reasonable to expect that rapid guessing percent correct rates may be higher (Wise & Ma, 2012).

**Response Time Effort (RTE)**

With the increase use of computers, one of the distinct advantages is to examine student effort, not only by the aggregate time a student spends on his or her test but also by the amount of time a student spends on each individual item. The initial introduction of response time effort (RTE) was made by Wise and Kong (2005), who hypothesized that students who lacked motivation would tend to answer questions quickly. In order to quantify the RTE for a student, a response time threshold was predetermined for all items. A crucial finding by Wise and Kong (2005) was that the RTE values of students were unrelated to ability. This result was essential to address the concern of a potentially confounding issue between student achievement and RTE.
In addition, the Wise and Kong (2005) study suggested that RTE values were better measures of test-taker effort than self-reported effort measures.

The response time threshold was applied to each item in order to determine if the student demonstrated sufficient effort, called solution behavior, on each item. When the response time was less than the response time threshold, it was deemed to be rapid-guessing behavior, which was classified as being noneffortful. The RTE represented a proportion of the solution, or effortful, behaviors (Wise & Kong, 2005). Values of 1.0 indicated strong motivation by the test-taker, while lesser values indicated less motivation by the test-taker. Each item that a student demonstrated solution behavior on was assigned a value of 1; each item that a student did not demonstrate solution behavior on, but rather rapid-guessing behavior, was assigned a value of 0. The sum of all values were added together and divided by the number of test items to determine a student’s RTE. For example, a student with an RTE of .80 demonstrated rapid-guessing behavior on 20% of the test items.

It is worth noting that in subsequent studies of response time effort (RTE), factors such as grade level, gender, content area, and the time of day when students were tested were shown to have impacted RTE data (Wise, Ma, Kingsbury, & Hauser, 2010). Student RTE declined beginning in grade 6, and the decline continued through middle school grade levels with an even greater decline in grade 9. Males also had lower mean effort than females; the RTE of males was also lower than females in both math and reading at every grade level (Wise et al., 2010). Between math and reading, lower RTE values were consistently found in reading when compared to math RTE values. Student effort also declined over the course of the day with mean effort being highest in the morning, declined at an accelerated pace as the day progressed, and
was its lowest at the end of the school day. However, the day of the week was not found to have an impact on student effort (Wise et al., 2010).

Initial research by Wise and Kong (2005) proposed a Common 3 threshold, which stated that a response time greater than or equal to three seconds for each item was deemed to be effortful or solution behavior while response time less than three seconds for each item was deemed to be noneffortful or rapid-guessing behavior. Recent research, however, has explored a variety of strategies to set response time thresholds, including bimodal distributions (Setzer et al., 2013) as well as using normative thresholds (NT) (Wise & Ma, 2012). The purpose of response time thresholds is to be able to distinguish between solution behavior and rapid-guessing behavior.

The selection of response time thresholds is somewhat arbitrary, does not always take into account various test item characteristics, and, thus, common thresholds may not always be appropriate for all items (Setzer et al., 2013). Setzer et al. (2013) proposed using bimodal distributions to determine response time thresholds, which may vary, for individual items. An additional response time threshold strategy, developed by Wise and Ma (2012), explored the use of normative thresholds (NT). The thresholds were determined by percentages of the mean response times for each individual item in a set of test item data. The NT percentages examined were 10%, 15%, and 20%. To set a response time threshold, the mean response time for a given item was multiplied by each percentage (10%, 15%, and 20%). It was determined that among the Common 3 method and the 10%, 15%, and 20% threshold methods that the 10% NT (NT10) was recommended. Although the Common 3 method was deemed to be inferior to the NT10 threshold, the Common 3 provided a good balance between two competing response time principles: a) accurately identifying as many legitimate instances of rapid-guessing behaviors as
possible while simultaneously b) avoiding classifying solution behaviors as being rapid-guessing behaviors (Wise & Ma, 2012).

The bimodal distribution and normative threshold approaches for determining appropriate response time thresholds may be used with both fixed form assessments and computerized adaptive tests; however, the use of either the bimodal distribution or the normative threshold approach with a computerized adaptive test presents challenges, since students taking computerized adaptive tests will see a wide variety of test items, depending on the ability of each student.

**Identifying Students with Test Anxiety**

Hill (Lowe et al., 2011) proposed that students began to experience test anxiety as early as the second grade. This finding reinforced a previous finding that suggested students began experiencing test anxiety as early as seven years of age (Connor, 2003). In his meta-analysis of test anxiety, Hembree (1988) noted that the prevalence of test anxiety increased sharply and began to have a negative effect on student performance between grades three and five before the test anxiety leveled off in the middle school grade levels and remained constant through high school before test anxiety declined at the post-secondary level. Given the fact that much of test anxiety research has focused on post-secondary students, it is essential to more closely examine the test anxiety experience of students in grades K-12. Furthermore, test anxiety research has taken more of a deductive approach, which has led to a limited understanding of what students themselves perceived to be the sources or causes of test anxiety (Bonaccio & Reeve, 2010).

The distribution of students who experienced test anxiety when comparing the test anxiety levels of students taking classroom assessments to the test anxiety levels of students taking NCLB assessments has also been shown to differ. However, while students with low or
moderate levels of classroom assessment test anxiety showed an increase in test anxiety levels when taking an NCLB assessment, the proportion of students with high levels of test anxiety did not show a statistically significant difference between taking the classroom assessment and the NCLB assessment (Segool et al., 2013). Furthermore, it was noted that students experienced higher levels of cognitive and physiological symptoms of test anxiety in a high-stakes testing setting than in a classroom testing setting. This was contrary to results found in the United Kingdom such that students experienced higher levels of test anxiety on low-stakes assessments than on medium- to high-stakes assessments (Putwain, 2008b).

One plausible explanation expressed by Segool et al. (2013) for Putwain’s finding was that the students in Putwain’s study who took the low- and moderate-stakes assessments completed a mock assessment that was still considered to be a “novel experience” (p. 496). This was in contrast to the typical low-stakes classroom testing situation that students have more experience with. The novelty of and the emphasis placed on NCLB assessments could potentially be contributing factors in increased test anxiety levels in younger students. Nonetheless, the discrepancy between the findings was an indicator that further research is needed to examine the relationship between test anxiety levels and NCLB-related assessments.

Parents were found to be a factor in the test anxiety levels of post-secondary students such that parental involvement was a reasonable predictor of student test anxiety and that a positive relationship was found between the academic education of parents and students’ test anxiety levels (Shadach & Ganor-Miller, 2013). In families where at least one parent had earned a college degree, parents were perceived by their children to be more involved than by students in families where neither parent had earned a college degree. The study also differentiated between parental attitudes and parental behaviors, which, interestingly enough, deemed that
parents’ behavior, such as parental involvement and overprotection, “not only mediate the effects of parental attitudes on TA (test anxiety), but were dominant enough to explain TA scores by themselves” (Shadach & Ganor-Miller, p. 593).

An additional finding by Shadach and Ganor-Miller (2013) was that parental involvement correlated to students’ test anxiety and worry but not emotionality, while students’ attitudes correlated to test anxiety and emotionality but not with worry. The parental pressure students experienced was shown to be directly associated with worry and test-irrelevant thinking, both components of test anxiety (Putwain, Woods, & Symes, 2010b). Gaining knowledge of what students themselves perceive to be contributing factors to the test anxiety they experience will provide information helpful in identifying and developing effective intervention strategies designed to reduce test anxiety and its debilitating effects on student achievement.

Although much is known about the test anxiety construct, given the wide range in the estimates of test anxiety, it is apparent that measurement of students’ test anxiety levels is limited. Norms have not been established to measure test anxiety levels. Due to the lack of cut point norms in test anxiety research and, thus, no systemic criteria being used to identify students as having low, moderate, or high debilitating test anxiety levels, widely varying estimates of test anxiety prevalence have been reported by researchers (Friedman & Bendas-Jacob, 1997). One exception is that the Test Anxiety Inventory (TAI) has established cut points; however, those norms were established decades ago and were based on undergraduate students (Spielberger, 1980). It was recently found that the TAI norms are out-of-date and, based on recent research, the TAI should be renormed as college undergraduate demographics have changed a great deal since 1980 (Szafranski et al., 2012). None of the other scale development measures to date have established cut points to classify students according to test anxiety levels, which would likely be
hastened if test anxiety surveys would become more commonly used in education (von der Embse et al., 2014).

Some studies have incorporated arbitrary cut points based on standard deviations or percentages to classify participants as having high or low levels of debilitating test anxiety. For example, Peleg (2009) identified students in the highest 25% of a test anxiety study as having high test anxiety and students in the bottom 25% as having low test anxiety. The concern with such an arbitrary approach is that it is based solely on the study’s participants rather than on a well-established statistically sound set of cut points created from a normally distributed population.

Teachers have reported observing higher levels of test anxiety in students on NCLB assessments than on classroom assessments, and teachers have also indicated experiencing higher levels of test anxiety themselves about NCLB assessments (Segool et al., 2013). The extent of teacher anxiety was found to have a strong correlation to student test anxiety levels (Hembree, 1988). Furthermore, the (math) anxiety of elementary teachers themselves was likely to be transferred to students (Wood, 1988). These findings, coupled with implications for future research from previous studies, which indicate a need for additional research to examine test anxiety levels of all students, but particularly the test anxiety levels of students receiving English Language Learner services, special education services, and by gender, provide ample opportunities for more refined test anxiety research.

In order to glean more accurate estimates in the prevalence of test anxiety in K-12 students in the United States, a more comprehensive, intentional approach of assessing students for test anxiety is necessary. Prior to addressing the group and individual needs of students with elevated levels of test anxiety, it is crucial to know which students are experiencing debilitating
effects of test anxiety. Given the dynamic nature of the relationship between test anxiety and student achievement as well as the diverse and unique characteristics of student groups and of individual students, a complexity exists regarding the relationships among such a broad array of factors. A key first step is to identify individual students and student groups who are most at risk for having high levels of debilitating test anxiety. This information will provide insightful information to researchers and practitioners which will aid in designing effective intervention strategies in an effort to reduce the effects of debilitating test anxiety.

Early test anxiety research, being unidimensional in nature, sought to distinguish between the presence and the absence of debilitating test anxiety, which is anxiety associated with negative student achievement. Within the first decade of test anxiety research, however, a multidimensional view of the test anxiety construct began to emerge. Facilitating anxiety, which was shown to enhance student achievement, became a new focal point of multidimensional test anxiety research as well as identifying new components of debilitating test anxiety. Debilitating test anxiety has been studied far more often than facilitating test anxiety (Hembree, 1988). In addition, Friedman and Bendas-Jacob (1997) demonstrated that social humiliation also contributed to the test anxiety that students experienced.

**Debilitating test anxiety.** This type of test anxiety was defined as test anxiety that has a negative impact on student achievement. Wine’s (1971) cognitive-attentional (interference) model described this type of test anxiety as the task-irrelevant thoughts students experience that interfere with their ability to remain focused during testing in such a way that it impaired the students cognitively, resulting in lower student achievement. It was estimated that 10% of students experienced test anxiety that impaired student achievement (Segool et al., 2013).
Facilitating test anxiety. This type of test anxiety was described as an “initial nervousness or tension that improves executive functioning and performance” (Lowe & Lee, 2008, p. 234). This anxiety occurs when moderate forms of test anxiety optimally enhance student performance. Facilitating test anxiety was first identified through the bidirectional theory work of Alpert and Haber (1960), who demonstrated that this type of anxiety is positively associated with students’ academic achievement.

Test Anxiety Scales

The test anxiety construct has been studied for over 60 years (Liebert & Morris, 1967; Mandler & Sarason, 1952). During that time it evolved from a unidimensional construct to a more complex, multidimensional construct. There are several test anxiety measurement scales, including many that are available online (Salend, 2011a). Some of the early test anxiety inventories included the Test Anxiety Questionnaire (TAQ) (Mandler & Sarason, 1952), the Test Anxiety Scale for Children (TASC) (Sarason, Davidson, Lighthall, Waite, & Ruebush, 1960), and the Test Anxiety Scale (TAS), which was also developed by Sarason in 1975. The TASC only measured a single dimension and was most appropriate for use with students in grades 1-6. Furthermore, the TASC is considered outdated due to the complex working of items; its outdated domain definition as the measurement of test anxiety has not kept pace with the conceptualization of test anxiety and is single dimensional (Wren & Benson, 2004).

As the test anxiety construct came to be viewed as multidimensional, new test anxiety measurement scales emerged which drew upon, refined, and revised items from previous test anxiety scales. These scales include the Worry Emotionality Questionnaire (Liebert & Morris, 1967) and the Test Anxiety Inventory (TAI) (Spielberger, 1980). Recent research has led to the development of more multidimensional scales, such as the Children’s Test Anxiety Scale
(CTAS) (Wren & Benson, 2004), Westside Test Anxiety Scale (Driscoll, 2004), the Test Anxiety Scale for Elementary Students (TAS-E) (Lowe & Ang, 2012; Lowe et al., 2011), and the Test Anxiety Inventory for Children and Adolescents (TAICA) (Lowe & Lee, 2008; Lowe et al., 2008).

In support of multidimensional test anxiety constructs, some research studies have utilized factor analyses which indicated the existence of four test anxiety factors (Lowe & Lee, 2005; Sarason, 1984). Sarason’s (1984) four factor dimensions were bodily symptoms, tension, test-irrelevant thinking, and worry. The four factor dimensions identified by Lowe and Lee (2005) were classified as a collection of debilitating test anxiety factors, which were cognitive obstruction/inattention, physiological hyperarousal, social humiliation, and worry. A very comparable factor structure was also confirmed for the Test Anxiety Scale for Elementary Students (TAS-E) (Lowe et al., 2011).

**Children’s Test Anxiety Scale (CTAS).** The CTAS is a 30-item self-reporting multidimensional measure of test anxiety. A four-point Likert scale was used with a range from Almost Never to Almost Always. Responses were as follows: Almost Never = 1, Some of the Time = 2, Most of the Time = 3, and Almost Always = 4. Thus, the total anxiety, as measured by CTAS, ranged from a score of 30 to 120. The three test anxiety CTAS subscales are Thoughts, Off-Task Behaviors, and Autonomic Reactions. The Thoughts subscale is comprised of 13 items; the Off-Task Behaviors subscale is comprised of eight items; and the Autonomic Reactions subscale is comprised of nine items (Wren & Benson, 2004). Thus, the subscale ranges for Thoughts, Off-Task Behaviors, and the Autonomic Reactions are 13 to 52, 8 to 32, and 9 to 36, respectively.
The Thoughts component includes the cognitive component of the test anxiety experienced by children; however, it is seen as being more encompassing than the worry component of adult test anxiety. Worry and test-irrelevant thoughts were combined as Wren and Benson (2004) hypothesized that students are less likely to differentiate between those two components than adults would. The Autonomic Reactions component was theorized due to children referring more often to somatic response to test-related stress than merely to emotions. Children’s dependency needs suggested a need for a behavior component of test anxiety, Off-Task Behaviors. This component includes the distracting behaviors and nervous habits observed in students experiencing test anxiety (Wren & Benson, 2004). An interesting note regarding the CTAS is that the term test anxiety was not used anywhere in the instrument. Instead, the title of the CTAS survey was “Test Attitude Survey” (Wren & Benson, 2004, p. 232).

The Cronbach reliability estimate of the development sample for the 30-item CTAS was .92. The Cronbach reliability estimates for the subscales were .89 for Thoughts, .85 for Autonomic Reactions, and .78 for Off-Task Behaviors. Internal consistency of the validation sample was .92 overall, with .89 for Thoughts, .82 for Autonomic Reactions, and .76 for Off-Task Behaviors. Grade level, gender, and race subsample Cronbach alphas ranged from .61 to .93 with all but two (grade 3 Off-Task Behaviors/.61; Asians Off-Task Behaviors/.68) of the subsamples greater than .70. Two-thirds of the subsample alphas were greater than .80 (Wren & Benson, 2004).

Two other multidimensional test anxiety scales were considered for the study, however, neither was yet commercially available at the time this study was conducted. The Test Anxiety Inventory for Children and Adolescents (TAICA) is a 45-item self-reporting multidimensional measure of test anxiety, which is used with students in grades 4-12. A 5-point Likert scale is
used with a range from 1 (never true) to 5 (always true). It consists of four debilitating test anxiety factors, which comprise the Total Anxiety Scale by taking the sum of each debilitating test anxiety subscale. The four debilitating test anxiety factors are Cognitive Obstruction/Inattention, Physiological Hyperarousal, Social Humiliation, and Worry (Lowe & Lee, 2005).

Seven items comprise the Cognitive Obstruction/Inattention subscale, which measures memory and attention difficulties; seven items comprise the Physiological Hyperarousal subscale, which measures the physical symptoms that accompany an evaluative situation; 10 items comprise the Social Humiliation subscale, which is a measure of the fear of failing a test or being criticized by others for how one performs on a test; and seven items comprise the Worry subscale, which measures the adverse effects on test performance that are caused by negative thinking (Lowe & Lee, 2008).

There are two additional factors associated with test anxiety but not defined as being debilitating factors. These factors are Performance Enhancement/Facilitation Anxiety, which enhances student achievement, and the Lie scale, which is a validity construct for TAICA designed to measure “ideal test behavior” (Lowe & Lee, 2008, p. 235). Each of these subscales is comprised of seven items on the TAICA survey. It should be noted that cut points have not yet been established for the TAICA to assist practitioners in determining what constitutes high, middle, and low levels of test anxiety (Whitaker Sena et al., 2007). In addition, a Spanish version of TAICA was also developed (Unruh & Lowe, 2010). At the time of this study, however, neither the TAICA nor the TAICA-Spanish was available for commercial use.

The Test Anxiety for Elementary Students (TAS-E) is another multidimensional measure of test anxiety and is a complement to the TAICA. It is designed for use with students in grades
two through six and consists of 30 items which are measured across four test anxiety subscales. The Total Test Anxiety Scale is the sum of the four test anxiety subscales, which include Physiological Hyperarousal, Social Concerns, Task Irrelevant Behavior, and Worry (Lowe et al., 2011).

Nine items comprise the Physiological Hyperarousal subscale, which measures the physical symptoms that accompany test anxiety. Eight items comprise the Task-Irrelevant Behavior subscale, which measures students’ restless behaviors that accompany test anxiety. Six items comprise the Social Concerns subscale, which measures students’ concerns about how students perceive how others view them when they do not do well on tests. Seven items comprise the Worry subscale, which measures the worrisome thoughts students have that are associated with testing. Furthermore, the developers of TAS-E included some cognitive interference items as part of the Worry subscale. This was due to younger students’ struggles to differentiate between worry and task-irrelevant thoughts (Lowe et al., 2011).

Summary

The literature provided a glimpse into a research history of the test anxiety construct, the evolution of quantifying student motivation, and test anxiety interventions. The increase in high-stakes testing combined with the increase in the prevalence of test anxiety in K-12 students in the United States has also created a sense of urgency not only for identifying which students are most susceptible to the debilitating effects of test anxiety but also for designing effective test anxiety intervention programs. It can best be summed up by von der Embse and Hasson (2012), who stated that it was “essential to address any variable which could improve student test scores” (p. 184). The objective of this literature review was to synthesize the theoretical frameworks of test anxiety with aggregate response time and response time effort (RTE) in an attempt to
explore relationships between those variables. Early test anxiety research conceptualized the test anxiety construct as being unidimensional (Mandler & Sarason, 1952). A shift in the test anxiety conceptualization occurred in the 1960s when work by Alpert and Haber (1960) and Liebert and Morris (1967) proposed that test anxiety consisted of multiple components and, thus, was multidimensional. Hembree’s (1988) meta-analysis found that worry, the cognitive component of test anxiety, was more strongly correlated to students’ test performances than emotionality.

A review of the literature reflected that anxiety may be manifested in a variety of ways, such as behaviorally, cognitively, and physiologically, which can be evident by a variety of primary characteristics (Huberty, 2010; Lowe et al., 2008; Wren & Benson, 2004). Regardless of how anxiety is manifested, a negative association between anxiety and academic achievement has been well established (Ergene, 2003; Hembree, 1988; Owens et al., 2008; Peleg, 2009; Rezazadeh & Tavakoli, 2009; von der Embse et al., 2013)
CHAPTER THREE: METHODS

Design

A nonexperimental correlational research design was used for this study to examine relationships between test anxiety and the response times of students taking a computerized adaptive math test. The test anxiety scale includes three subscales, which were examined for relationships with the response times. The response times were categorized as aggregate, which is total response time, and response time effort (RTE), which is the proportion, represented as a decimal value, of test items for which a student exhibited solution behavior. The research design method used for this study is identified below, including research questions, research hypotheses, participants, setting, instrumentation, procedures, and data analysis.

A research gap of the relationships between test anxiety and response time exists. A nonexperimental correlational design for this study was appropriate given that it was helpful to determine the degree and direction of the relationship between test anxiety and the test anxiety subscales with aggregate response time and response time effort (RTE). Two distinct advantages of a correlational design are the ability to analyze the relationship among multiple variables, as well as the ability to provide information about the degree of the relationship between the variables (Gall, Gall, & Borg, 2007). It must be noted, however, that making causal inferences about the statistically significant relationships found in a correlational study would be inappropriate (Gall et al., 2007).

The nonexperimental correlational research design for this study examined various relationships between test anxiety and each of the three subscales with each of the two types of response times, aggregate and response time effort (RTE), on a cognitive adaptive test, the Measures of Academic Progress® (MAP®) math test by the Northwest Evaluation Association™
(NWEA™). A multidimensional four-point Likert test anxiety scale, the Children’s Test Anxiety Scale (CTAS), was used to determine each participant’s total test anxiety as well as his or her test anxiety level for each subscale. The three test anxiety subscales of the CTAS are Thoughts, Off-Task Behaviors, and Autonomic Reactions (Wren & Benson, 2004).

Response time was calculated in two manners. The aggregate response time is the total time participants spend taking the MAP® math test, while the response time effort (RTE) is a sum of dichotomized values that are based on an a post-hoc calculated threshold to indicate a level of student motivation. The primary difference between aggregate response time and RTE is that RTE has been shown to provide more dynamic insight into student effort. Response time effort (RTE) takes into account the effort, measured to the tenths of a second, that a student puts forth on each individual item of an assessment (Wise & Kong, 2005). Aggregate response time merely determines the total amount of a time a student spends on an entire assessment without regard to a student’s effort on each individual item.

The correlations examined in this study were between each of the four test anxiety scales with each response time measure, the aggregate response time and the response time effort (RTE) (see Figure 1). In addition, the variable of interest in this study that was controlled for is gender. Examining the correlation between test anxiety and each of the response time measures, aggregate and RTE, revealed which type of response time measure had a potentially stronger correlation with test anxiety and also with which particular test anxiety subscale. It was anticipated that test anxiety has a stronger correlation with RTE; however, RTE data availability is not as prevalent as aggregate response time data. Determining the significance of a correlation between test anxiety and RTE was helpful in determining how crucial it may be for practitioners to obtain such data.
<table>
<thead>
<tr>
<th></th>
<th>Total Test Anxiety</th>
<th>Thoughts</th>
<th>Off-Task Behaviors</th>
<th>Autonomic Reactions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total Test Anxiety</td>
<td>Thoughts</td>
<td>Off-Task Behaviors</td>
<td>Autonomic Reactions</td>
</tr>
<tr>
<td>Aggregate Response Time</td>
<td>Aggregate Response Time – off-Task Behavior</td>
<td>Aggregate Response Time – Autonomic Reactions</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Figure 1.** Test Anxiety Response Time Correlations. Total Test Anxiety is the sum of the three Children’s Test Anxiety Survey (CTAS) subscales: a) Thoughts, b) Off-Task Behaviors, and c) Autonomic Reactions (Wren & Benson, 2004). Aggregate Response Time – the total time, or duration, a student spends to complete a test. Response Time Effort (RTE) – the proportion, represented as a decimal value, of test items for which a student exhibited solution behavior (Wise & Kong, 2005). Correlations were also examined by gender.

**Research Questions**

For the research questions, the four measures of test anxiety consisted of the three subscales of the Children’s Test Anxiety Scale (CTAS), Thoughts, Off-Task Behavior, and Autonomic Reactions, as well as the total test anxiety, which is a sum of the three CTAS subscales. The research questions for this nonexperimental correlational study were as follows:
**RQ1:** Is there a relationship between any of the four measures of test anxiety as measured by the Children’s Test Anxiety Scale (CTAS) and the response time effort (RTE) of students taking the Measures of Academic Progress® (MAP®) math assessment after separately controlling for gender?

**RQ2:** Is there a relationship between any of the four measures of test anxiety as measured by the Children’s Test Anxiety Scale (CTAS) and the aggregate response time of students taking the Measures of Academic Progress® (MAP®) math assessment after separately controlling for gender?

**Null Hypotheses**

Sixteen hypotheses were examined as part of the study. The null hypotheses for the first research question, which pertained to relationships between the four measures of test anxiety as measured by the Children’s Test Anxiety Scale (CTAS) survey and the seventh grade math students’ response time effort (RTE) while taking the Northwest Evaluation Association’s™ Measures of Academic Progress® (MAP®) math assessment, in the study were as follows:

**H01:** There is no statistically significant relationship between students’ **Total Test Anxiety** as measured by the Children’s Test Anxiety Scale (CTAS) and students’ **response time effort** (RTE) while taking the Measures of Academic Progress® (MAP®) math assessment.

**H02:** There is no statistically significant relationship between students’ **Total Test Anxiety** as measured by the Children’s Test Anxiety Scale (CTAS) and students’ **response time effort** (RTE) while taking the Measures of Academic Progress® (MAP®) math assessment after controlling for gender.
H₀³ There is no statistically significant relationship between the **Thoughts** subscale as measured by the Children’s Test Anxiety Scale (CTAS) and students’ **response time effort** (RTE) while taking the Measures of Academic Progress® (MAP®) math assessment.

H₀⁴: There is no statistically significant relationship between the **Thoughts** subscale as measured by the Children’s Test Anxiety Scale (CTAS) and students’ **response time effort** (RTE) while taking the Measures of Academic Progress® (MAP®) math assessment after controlling for gender.

H₀⁵: There is no statistically significant relationship between the **Off-Task Behaviors** subscale as measured by the Children’s Test Anxiety Scale (CTAS) and students’ **response time effort** (RTE) while taking the Measures of Academic Progress® (MAP®) math assessment.

H₀⁶: There is no statistically significant relationship between the **Off-Task Behaviors** subscale as measured by the Children’s Test Anxiety Scale (CTAS) and students’ **response time effort** (RTE) while taking the Measures of Academic Progress® (MAP®) math assessment after controlling for gender.

H₀⁷: There is no statistically significant relationship between the **Autonomic Reactions** subscale as measured by the Children’s Test Anxiety Scale (CTAS) and students’ **response time effort** (RTE) while taking the Measures of Academic Progress® (MAP®) math assessment.

H₀⁸: There is no statistically significant relationship between the **Autonomic Reactions** subscale as measured by the Children’s Test Anxiety Scale (CTAS) and students’
response time effort (RTE) while taking the Measures of Academic Progress® (MAP®) math assessment after controlling for gender.

The null hypotheses for the second research question, which pertained to relationships between the students’ Total Test Anxiety as measured by the Children’s Test Anxiety Scale (CTAS) survey and the students’ aggregate response time while taking the Northwest Evaluation Association’s™ Measures of Academic Progress® (MAP®) math assessment, in the study were as follows:

**H₀9:** There is no statistically significant relationship between students’ Total Test Anxiety as measured by the Children’s Test Anxiety Scale (CTAS) and students’ aggregate response time while taking the Measures of Academic Progress® (MAP®) math assessment.

**H₀10:** There is no statistically significant relationship between students’ Total Test Anxiety as measured by the Children’s Test Anxiety Scale (CTAS) and students’ aggregate response time while taking the Measures of Academic Progress® (MAP®) math assessment after controlling for gender.

**H₀11:** There is no statistically significant relationship between the Thoughts subscale as measured by the Children’s Test Anxiety Scale (CTAS) and students’ aggregate response time while taking the Measures of Academic Progress® (MAP®) math assessment.

**H₀12:** There is no statistically significant relationship between the Thoughts subscale as measured by the Children’s Test Anxiety Scale (CTAS) and students’ aggregate response time while taking the Measures of Academic Progress® (MAP®) math assessment after controlling for gender.
**H₀13:** There is no statistically significant relationship between the **Off-Task Behaviors** subscale as measured by the Children’s Test Anxiety Scale (CTAS) and students’ **aggregate response time** while taking the Measures of Academic Progress® (MAP®) math assessment.

**H₀14:** There is no statistically significant relationship between the **Off-Task Behaviors** subscale as measured by the Children’s Test Anxiety Scale (CTAS) and students’ **aggregate response time** while taking the Measures of Academic Progress® (MAP®) math assessment after controlling for gender.

**H₀15:** There is no statistically significant relationship between the **Autonomic Reactions** subscale as measured by the Children’s Test Anxiety Scale (CTAS) and students’ **aggregate response time** while taking the Measures of Academic Progress® (MAP®) math assessment.

**H₀16:** There is no statistically significant relationship between the **Autonomic Reactions** subscale as measured by the Children’s Test Anxiety Scale (CTAS) and students’ **aggregate response time** while taking the Measures of Academic Progress® (MAP®) math assessment after controlling for gender.

**Participants and Setting**

A nonrandom convenience sample of students in seventh grade math classes was used in this study. There were 215 of an estimated 400 eligible students who participated in the study. The sample consisted of 99 females and 116 males. There were 177 Caucasian students, 20 Hispanic students, nine multiracial students, and eight Asian students who participated in the study. Student ages ranged from 11 to 13. Students in grade seven were chosen as a result of the literature review, which indicated that middle school students typically experience higher test
anxiety levels than elementary students. Students with disabilities who, according to their Individualized Education Plan (IEP), are exempt from general or modified tests, such as state and national assessments as well as cognitive adaptive tests like the NWEA™ MAP® math test, were not included in the non-random convenience sample.

For a Pearson product-moment zero-order correlation based on a level of significance of \( \alpha = .05 \) and a statistical power of .7 with a medium effect size, the minimum sample size needed was 66 (Olejnik, 1984; Warner, 2013). For the correlation of the gender subgroup in this study and based on a level of significance of \( \alpha = .05 \) and a statistical power of .7 with a medium effect size, the minimum sample size needed for such a subgroup was 44 (Olejnik, 1984; Warner, 2013). Given the combination of needed sample sizes for the zero-order and subsample correlations, which would be for the gender subgroup, a minimum of 88 participants was needed provided that the 44 participant minimum was also met for each subgroup category (female and male). The required minimums were met (\( N = 215; \) females = 99; males = 116).

The setting was in a middle school in a large rural school district in the West with an estimated districtwide K-12 population of 6000. The school district has five K-6 elementary schools, one middle school, and two high schools. The demographics of the participating middle school site include an estimated student population of 800 that is 0.9% African-American, and 0.7% American Indian, 3.4% Asian, 82.2% Caucasian, 8.9% Hispanic, and 3.7% multiracial. Schoolwide, 47.2% of the students are females while 52.8% of the students are males. Approximately 9% of the students in the middle school qualify for the free and reduced lunch program.

Participants took the NWEA™ MAP® math test in one of three computers labs throughout the middle school. Students were also tested during the morning and during the afternoon
depending on when their class was scheduled to take the NWEA™ MAP® math test. The groups in the study were intact groups and, thus, assignment to the groups in the study was non-random, as students were grouped according to their academic excellence (homeroom) teacher. The size of each intact group was between 25-30 students. Classroom teachers proctored the NWEA™ MAP® math test and facilitated the Children’s Test Anxiety Scale (CTAS) survey immediately after a participant in the study finished the NWEA™ MAP® math test.

**Instrumentation**

The instrument used in this study was the Children’s Test Anxiety Scale (CTAS), which is a four-point Likert scale 30-item self-report survey which examines multidimensions of test anxiety (Wren & Benson, 2004). Many of the early test anxiety surveys measured one dimension of test anxiety or used high school or college populations to develop test anxiety surveys (Hembree, 1988). Hembree’s meta-analysis also noted that the prevalence of test anxiety increased in grades 3-5. In addition, the wording of some test anxiety items were outdated, domain definitions were outdated, and there were issues with dimensionality (Wren & Benson, 2004). This led to test anxiety research with emphases on the development of test anxiety surveys that were multidimensional and the development of test anxiety surveys to be used with elementary student populations. The CTAS instrument addresses both of those areas of emphasis and has been used in numerous studies (Cassady & Finch, 2014; Nyroos, Korhonen, Linnanmäki, & Svens-Liavág, 2012; Segool et al., 2013).

The purpose of this instrument was to measure multiple dimensions of test anxiety, which consists of Total Test Anxiety and three distinct test anxiety subscales: Thoughts, Off-Task Behaviors, and Autonomic Reactions. Thoughts are the worry cognitions that students experience during testing, such as self-critical thoughts, test-related concerns, and test-irrelevant
thoughts. Off-Task Behaviors are the nervous habits and distracting behaviors that can be observed of students during testing, such as rocking, tapping one’s feet, playing with pencils, and any behaviors deemed to be inattentive or distracting. Autonomic Reactions are the physiological symptoms a student experiences during testing, such as perspiration, stomach problems, and headaches, for example (Wren & Benson, 2004).

A development sample (N = 230) of students in grades 3-6 was used to obtain item analysis and reliability estimates for what was initially a 50-item CTAS scale. The CTAS items were written for students with at least a grade 3 reading level. This resulted in a refined 30-item CTAS scale, which was then administered to a validation sample (N = 261) of students in grades 3-6 in order to obtain construct validity evidence. A 25-item CTAS scale was created in an effort to make the CTAS scale more efficient (See Appendix A for Children’s Test Anxiety Scale). Although the 25-item CTAS scale did reveal a statistically significant chi-square, the authors felt that the 25-item scale did not vary appreciably from the 30-item scale (Wren & Benson, 2004).

The Thoughts subscale is comprised of 13 items; the Autonomic Reactions subscale is comprised of nine items; and the Off-Task Behaviors subscale is comprised of eight items (Wren & Benson, 2004). When summed together the three CTAS subscales provide a Total Test Anxiety index value. The instrument used a four-point Likert scale with a range from Almost Never to Almost Always. Responses were as follows: Almost Never = 1, Some of the Time = 2, Most of the Time = 3, and Almost Always = 4. Thus, the range of the Thoughts subscale is 13 to 52 where 13 is the lowest possible score meaning that the participant’s Thoughts test anxiety is at a minimum, while 52 is the highest possible score meaning that the participant’s Thoughts test anxiety is at a maximum. In addition, the range of the Autonomic Reactions subscale is 9 to 36.
where 9 is the lowest possible score meaning that the participant’s Autonomic Reactions test anxiety is at a minimum, while 36 is the highest possible score meaning that the participant’s Autonomic Reactions test anxiety is at a maximum. The range of the Off-Task Behaviors subscale is 8 to 32 where 8 is the lowest possible score meaning that the participant’s Off-Task Behaviors test anxiety is at a minimum, while 32 is the highest possible score meaning that the participant’s Off-Task Behaviors test anxiety is at a maximum. The range of Total Test Anxiety is consequently 30 to 120 where 30 is the lowest possible score meaning that the participant’s Total Test Anxiety is at a minimum, while 120 is the highest possible score meaning that the participant’s Total Test Anxiety is at a maximum. The CTAS did not provide specific instructions on how best to facilitate the CTAS survey. According to Croyle et al. (2012), however it is best practice to measure student’s test anxiety immediately following an assessment. That was the protocol that was followed in this study.

The reliability estimate of the development sample for the 30-item version of the CTAS was .92. The reliability estimates for the subscales for Thoughts, Off-Task Behaviors, and Autonomic Reactions were .89, .78, and .85, respectively. Internal consistency of the validation sample was .92 overall. The internal consistency for the CTAS subscales for Thoughts, Off-Task Behaviors, and Autonomic Reactions were .89, .76, and .82, respectively. Grade level, gender, and race subsample alphas ranged from .61 to .93. All but two of the subsamples were greater than .70. The two subsamples with alpha values less than .70 were Grade 3 Off-Task Behaviors (.61) and Asian Off-Task Behaviors (.68). Two-thirds of the subsample alphas were greater than .80 (Wren & Benson, 2004).

The long form of the NWEA™ MAP® math test has 50 items and took students approximately 60 minutes to complete; however, it is an untimed test so the response times of
students vary. The CTAS survey took students 6-8 minutes to complete and was taken upon completion of the NWEA™ MAP® math test using the website Survey Monkey. The CTAS survey scores were gathered and, thus, scored electronically. The NWEA™ MAP® math test response time data were also captured electronically. Two types of response times, aggregate response time and response time effort, were used for this study. Aggregate response time is the total amount of time participants spend taking a test. Response time effort (RTE), a more dynamic measure of students’ motivation, is based on the proportion of students’ solution behaviors. Solution behaviors are distinguished from rapid-guessing behaviors by a predetermined time threshold for each item (Wise & Kong, 2005).

The RTE threshold for this study was 10% of the mean aggregate response time divided by the number of items on the NWEA™ MAP® math test participants take. Each item that a student demonstrated solution behavior on was assigned a value of 1; each item that a student did not demonstrate solution behavior on, but rather rapid-guessing behavior, was assigned a value of 0. The sum of all values were added together and divided by the number of test items to determine a student’s RTE. Values of 1.0 indicate strong motivation by the test-taker, while lesser values indicate less motivation by the test-taker. For example, a student with an RTE of .80 would have demonstrated rapid-guessing behavior on 20% of the test items. Permission to use the CTAS survey as part of this research study was granted by Dr. Douglas G. Wren (See Appendix C for the permission letter).

**Procedures**

A nonexperimental correlational research design was used for this study. The study received IRB approval (see Appendix D for the IRB approval letter). Students in seventh grade math classes were invited to participate in the study. Students in seventh grade math classrooms
were provided with an informed consent form for their parents or guardians to sign in order for the student to participate in the study (see Appendix E for the participant consent form).

Students were also provided with a child assent form (see Appendix F for the assent form). In addition, a recruitment letter was emailed to all parents and guardians of seventh grade students (see Appendix G for the recruitment letter). This letter provided parents and guardians with a thorough explanation of the study. Students’ confidentiality was protected throughout the study in accordance with Institutional Review Board (IRB) expectations, including the participants’ right to leave the study at any time.

Seventh grade students received a letter describing the study. An after school informational meeting was held to provide students and parents with an opportunity to ask questions, learn more about the study, and for the researcher to address any concerns anyone had about the study. A pilot study of the CTAS survey was conducted by the researcher with a student population of seventh grade math students not participating in the study. The pilot study helped gauge the amount of time it takes for participants to complete the survey, to ensure that the CTAS survey data was properly captured by Survey Monkey, and to determine ways to improve the survey process for both participants and those who were to administer the CTAS survey. In addition, two training sessions were conducted for classroom teachers who administered the NWEA™ MAP® math test and the CTAS survey to ensure that the testing and survey protocols were followed.

Participants took the Northwest Evaluation Association™ (NWEA™) Measures of Academic Progress® (MAP®) math test. Immediately following completion of the MAP® math test, participants took the Children’s Test Anxiety Scale (CTAS) survey online. The CTAS survey was administered using Survey Monkey. The same identification number was used by
each participant for both the NWEA™ MAP® math test and the CTAS survey in order to match data between the MAP® test and CTAS survey. NWEA™ offered and provided the researcher with the response time data in a Microsoft Excel spreadsheet within two weeks of the administration of the NWEA™ MAP® math test. To ensure the participants’ confidentiality, the participating district’s assessment department offered and collected the NWEA™ Math test response time data and CTAS survey data. Once the NWEA™ Math test response data and CTAS survey data were merged into one file, all school district-issued identification numbers were replaced by new numbers using a random number generator until each participant had a new identification number in order to further ensure confidentiality of the participants.

The researcher added response time threshold formulas to the MAP® math test data to identify which item responses were deemed effortful and which items were deemed to not be effortful in order to calculate the response time effort (RTE) for each participant. A normative threshold (NT) of 7.5 seconds was used in the study. The resulting data was then analyzed using SPSS software (IBM Corporation, 2015).

**Data Analysis**

This study utilized Spearman correlations to analyze relationships among the variables of interest: aggregate response time, response time effort, total test anxiety, Thoughts, Off-Task Behaviors, and Autonomic Reactions. Furthermore, in order to examine relationships among the variable of interest while controlling for the influence of gender, correlations by subsample for each gender were also used in this study (Gall et al., 2007).

Students took the Northwest Evaluation Association™ (NWEA™) Measures of Academic Progress® (MAP®) math test. Students’ aggregate response time and response time effort (RTE) data captured from the NWEA™ MAP® math test was used for this study. To calculate the RTE
values for each participant, the Common threshold for this study was first determined by finding
the mean aggregate response time for all participants and setting the RTE threshold at 10% of the
mean aggregate response time divided by the number of items on the MAP® math test
participants took.

Given that the response time data that NWEA™ provided to the researcher for the study
were rounded to the nearest second using standard rounding procedures, the threshold was set at
the nearest 0.5 second value, that is, 3.5 seconds, 4.5 seconds, and so on. Following the
completion of the NWEA™ MAP® math test, participants completed the Children’s Test Anxiety
Scale (CTAS) survey to determine participants’ Total Test Anxiety and Subscale Test Anxiety
levels. The CTAS subscales measured were Thoughts, Off-Task Behaviors, and Autonomic
Reactions.

The response time effort (RTE) value for each participant was a sum of the dichotomized
response time values for each participant’s items on the Northwest Evaluation Association™
(NWEA™) Measures of Academic Progress® (MAP®) math test. The RTE index value is a ratio
scale of measurement. In addition, the CTAS four-point Likert scale results were summed to
create four indices, one index for the total test anxiety which is a sum of the three CTAS
subscales ( Thoughts, Off-Task Behaviors, and Autonomic Reactions), as well as an index for
each of the CTAS subscales, thus, creating three latent variables of CTAS.

Pearson product-moment correlations will be used with Likert-scale indices, which has
been supported by previous studies (Jacqueline, 2013; Maurer & Pierce, 1998; Norman, 2010).
Bivariate data screening included creating scatterplots. Assumption tests included analyzing
scatterplots for the variables of interest in the study to ensure that normality assumptions were
not violated. The statistical analysis used for each hypotheses in the study was a = .05. The
correlations were interpreted in accordance with Cohen’s d (1988) conventions of .2, .5, and .8, which represent small, medium, and large effect sizes, respectively, which correspond to weak, moderate, and strong effect sizes for Spearman r conventions of .10, .24., and .37, respectively (Warner, 2013)

A Spearman correlation was used to examine each relationship that exists between the CTAS Total Test Anxiety and each of the CTAS subscales (Thoughts, Off-Task Behaviors, and Autonomic Reactions) with the students’ response time effort (RTE) on the Measures of Academic Progress® (MAP®) math assessment. The Spearman correlations between RTE and each of the test anxiety measures of CTAS Total Test Anxiety, Thoughts, Off-Task Behaviors, and Autonomic Reactions were examined with hypotheses one, three, five, and seven, respectively. A Spearman correlation was also used to examine each relationship that existed between the CTAS Total Test Anxiety and each of the CTAS subscales (Thoughts, Off-Task Behaviors, and Autonomic Reactions) with the students’ aggregate response time on the Measures of Academic Progress® (MAP®) math assessment. The Spearman correlations between aggregate response time and each of the test anxiety measures of CTAS Total Test Anxiety, Thoughts, Off-Task Behaviors, and Autonomic Reactions were examined with hypotheses nine, eleven, thirteen, and fifteen, respectively.

A Spearman correlation was used to examine each relationship, by gender, that exists between the CTAS Total Test Anxiety and each of the CTAS subscales (Thoughts, Off-Task Behaviors, and Autonomic Reactions) with the students’ response time effort (RTE) on the Measures of Academic Progress® (MAP®) math assessment. The correlations between RTE and each of the test anxiety measures of CTAS Total Test Anxiety, Thoughts, Off-Task Behaviors, and Autonomic Reactions were examined, by gender, with hypotheses two, four, six, and eight,
respectively. A correlation was also used to examine each relationship, by gender, that existed between the CTAS Total Test Anxiety and each of the CTAS subscales (Thoughts, Off-Task Behaviors, and Autonomic Reactions) with the students’ aggregate response time on the Measures of Academic Progress® (MAP®) math assessment. The correlations between aggregate response time and each of the test anxiety measures of CTAS Total Test Anxiety, Thoughts, Off-Task Behaviors, and Autonomic Reactions were examined, by gender, with hypotheses ten, twelve, fourteen, and sixteen, respectively.
CHAPTER FOUR: FINDINGS

Research Questions

**RQ1:** Is there a relationship between any of the four measures of test anxiety as measured by the Children’s Test Anxiety Scale (CTAS) and the response time effort (RTE) of seventh grade math students taking the Measures of Academic Progress® (MAP®) math assessment after separately controlling for gender?

**RQ2:** Is there a relationship between any of the four measures of test anxiety as measured by the Children’s Test Anxiety Scale (CTAS) and the aggregate response time of seventh grade math students taking the Measures of Academic Progress® (MAP®) math assessment after separately controlling for gender?

Null Hypotheses

Sixteen hypotheses were examined as part of the study. The null hypotheses for the first research question, which pertained to relationships between the four measures of test anxiety as measured by the Children’s Test Anxiety Scale (CTAS) survey and the students’ response time effort (RTE) while taking the Northwest Evaluation Association’s™ (NWEA™) Measures of Academic Progress® (MAP®) math assessment, in the study were as follows:

**H₀1:** There is no statistically significant relationship between students’ Total Test Anxiety as measured by the Children’s Test Anxiety Scale (CTAS) and students’ response time effort (RTE) while taking the Measures of Academic Progress® (MAP®) math assessment.

**H₀2:** There is no statistically significant relationship between students’ Total Test Anxiety as measured by the Children’s Test Anxiety Scale (CTAS) and students’ response
time effort (RTE) while taking the Measures of Academic Progress® (MAP®) math assessment after controlling for gender.

**H₀₃** There is no statistically significant relationship between the **Thoughts** subscale as measured by the Children’s Test Anxiety Scale (CTAS) and students’ **response time effort** (RTE) while taking the Measures of Academic Progress® (MAP®) math assessment.

**H₀₄:** There is no statistically significant relationship between the **Thoughts** subscale as measured by the Children’s Test Anxiety Scale (CTAS) and students’ **response time effort** (RTE) while taking the Measures of Academic Progress® (MAP®) math assessment after controlling for gender.

**H₀₅:** There is no statistically significant relationship between the **Off-Task Behaviors** subscale as measured by the Children’s Test Anxiety Scale (CTAS) and students’ **response time effort** (RTE) while taking the Measures of Academic Progress® (MAP®) math assessment.

**H₀₆:** There is no statistically significant relationship between the **Off-Task Behaviors** subscale as measured by the Children’s Test Anxiety Scale (CTAS) and students’ **response time effort** (RTE) while taking the Measures of Academic Progress® (MAP®) math assessment after controlling for gender.

**H₀₇:** There is no statistically significant relationship between the **Autonomic Reactions** subscale as measured by the Children’s Test Anxiety Scale (CTAS) and students’ **response time effort** (RTE) while taking the Measures of Academic Progress® (MAP®) math assessment.
H08: There is no statistically significant relationship between the Autonomic Reactions subscale as measured by the Children’s Test Anxiety Scale (CTAS) and students’ response time effort (RTE) while taking the Measures of Academic Progress\textsuperscript{\textregistered} (MAP\textsuperscript{\textregistered}) math assessment after controlling for gender.

The null hypotheses for the second research question, which pertained to relationships between the students’ Total Test Anxiety as measured by the Children’s Test Anxiety Scale (CTAS) survey and the students’ aggregate response time while taking the Northwest Evaluation Association’s\textsuperscript{\texttrademark} Measures of Academic Progress\textsuperscript{\textregistered} (MAP\textsuperscript{\textregistered}) math assessment, in the study were as follows:

H09: There is no statistically significant relationship between students’ Total Test Anxiety as measured by the Children’s Test Anxiety Scale (CTAS) and students’ aggregate response time while taking the Measures of Academic Progress\textsuperscript{\textregistered} (MAP\textsuperscript{\textregistered}) math assessment.

H010: There is no statistically significant relationship between students’ Total Test Anxiety as measured by the Children’s Test Anxiety Scale (CTAS) and students’ aggregate response time while taking the Measures of Academic Progress\textsuperscript{\textregistered} (MAP\textsuperscript{\textregistered}) math assessment after controlling for gender.

H011: There is no statistically significant relationship between the Thoughts subscale as measured by the Children’s Test Anxiety Scale (CTAS) and students’ aggregate response time while taking the Measures of Academic Progress\textsuperscript{\textregistered} (MAP\textsuperscript{\textregistered}) math assessment.

H012: There is no statistically significant relationship between the Thoughts subscale as measured by the Children’s Test Anxiety Scale (CTAS) and students’ aggregate
response time while taking the Measures of Academic Progress® (MAP®) math assessment after controlling for gender.

**H₀13:** There is no statistically significant relationship between the Off-Task Behaviors subscale as measured by the Children’s Test Anxiety Scale (CTAS) and students’ aggregate response time while taking the Measures of Academic Progress® (MAP®) math assessment.

**H₀14:** There is no statistically significant relationship between the Off-Task Behaviors subscale as measured by the Children’s Test Anxiety Scale (CTAS) and students’ aggregate response time while taking the Measures of Academic Progress® (MAP®) math assessment after controlling for gender.

**H₀15:** There is no statistically significant relationship between the Autonomic Reactions subscale as measured by the Children’s Test Anxiety Scale (CTAS) and students’ aggregate response time while taking the Measures of Academic Progress® (MAP®) math assessment.

**H₀16:** There is no statistically significant relationship between the Autonomic Reactions subscale as measured by the Children’s Test Anxiety Scale (CTAS) and students’ aggregate response time while taking the Measures of Academic Progress® (MAP®) math assessment after controlling for gender.

**Descriptive Statistics**

Both females (n = 99) and males (n = 116) participated in this study. The mean response time effort (RTE) value of all participants (N = 215) was 0.9928 (SD = .019) which suggests that participants demonstrated solution behavior on 99.28% of the NWEA™ MAP® math test items. The majority of the response time effort (RTE) scores were 1.00 (76.7%), which indicated that
based on the normative threshold, 7.5 seconds, used in the study that 76.7% of participants were
demed to have exhibited solution behavior on every test item on the NWEA™ MAP® math
test, with the second highest RTE value being .98 (15.8%). The mean aggregate response time
(ART) of all participants on the NWEA™ MAP® math test was 67.18 minutes (SD = 25.02).

Results

Data screening of the scatterplots indicated skewness of the data in the study. The
scatterplots revealed violations of the normality assumptions for using a Pearson product-
moment correlations in the study. Thus, Spearman correlations were used (Warner, 2013). The
statistical analysis used for each hypotheses in the study was the conventional α = .05. Cohen’s
d (1988) conventions were used for the Spearman correlations. Weak, moderate, and strong
effect sizes were used for Spearman r conventions of .10, .24., and .37, respectively (Warner, 2013).

Research question one asked, “Is there a relationship between any of the four measures of
test anxiety as measured by the Children’s Test Anxiety Scale (CTAS) and the response time
effort (RTE) of students taking the Measures of Academic Progress® (MAP®) math assessment
after separately controlling for gender?” There were a total of eight hypotheses related to this
research question numbered one through eight.

Null Hypothesis One

Null hypothesis one was, “There is no statistically significant relationship between
students’ Total Test Anxiety as measured by the Children’s Test Anxiety Scale (CTAS) and
students’ response time effort (RTE) while taking the Measures of Academic Progress® (MAP®)
math assessment.” A Spearman correlation was used because the normality assumption for a
Pearson correlation was violated. A weak negative relationship was found between Total Test Anxiety and RTE ($r_s = -.06, p = .40$). This finding failed to reject null hypothesis one.

**Null Hypothesis Two**

Null hypothesis two was, “There is no statistically significant relationship between students’ Total Test Anxiety as measured by the Children’s Test Anxiety Scale (CTAS) and students’ response time effort (RTE) while taking the Measures of Academic Progress® (MAP®) math assessment after controlling for gender.” A Spearman correlation was used because the normality assumption for a Pearson correlation was violated. No relationship was found in the female subsample between Total Test Anxiety and RTE ($r_s = .00, p = .99$). For the male subsample, a weak negative relationship was found between Total Test Anxiety and RTE ($r_s = -.12, p = .22$). These findings failed to reject null hypothesis two.

**Null Hypothesis Three**

Null hypothesis three was, “There is no statistically significant relationship between the Thoughts subscale as measured by the Children’s Test Anxiety Scale (CTAS) and students’ response time effort (RTE) while taking the Measures of Academic Progress® (MAP®) math assessment.” A Spearman correlation was used because the normality assumption for a Pearson correlation was violated. A weak negative relationship was found between Thoughts and RTE ($r_s = -.11, p = .13$). This finding failed to reject null hypothesis three.

**Null Hypothesis Four**

Null hypothesis four was, “There is no statistically significant relationship between the Thoughts subscale as measured by the Children’s Test Anxiety Scale (CTAS) and students’ response time effort (RTE) while taking the Measures of Academic Progress® (MAP®) math assessment after controlling for gender.” A Spearman correlation was used because the
normality assumption for a Pearson correlation was violated. A weak negative relationship was found in the female subsample between Thoughts and RTE ($r_s = -.07, p = .52$). For the male subsample, a weak negative relationship was found between Thoughts and RTE ($r_s = -.14, p = .14$). These findings failed to reject null hypothesis four.

**Null Hypothesis Five**

Null hypothesis five was, “There is no statistically significant relationship between the Off-Task Behaviors subscale as measured by the Children’s Test Anxiety Scale (CTAS) and students’ response time effort (RTE) while taking the Measures of Academic Progress® (MAP®) math assessment.” A Spearman correlation was used because the normality assumption for a Pearson correlation was violated. A very weak negative relationship was found between Off-Task Behaviors and RTE ($r_s = -.02, p = .76$). This finding failed to reject null hypothesis five.

**Null Hypothesis Six**

Null hypothesis six was, “There is no statistically significant relationship between the Off-Task Behaviors subscale as measured by the Children’s Test Anxiety Scale (CTAS) and students’ response time effort (RTE) while taking the Measures of Academic Progress® (MAP®) math assessment after controlling for gender.” A Spearman correlation was used because the normality assumption for a Pearson correlation was violated. A very weak negative relationship was found in the female subsample between Off-Task Behavior and RTE ($r_s = -.01, p = .95$). For the male subsample, a very weak negative relationship was found between Off-Task Behavior and RTE ($r_s = -.04, p = .66$). These findings failed to reject null hypothesis six.

**Null Hypothesis Seven**

Null hypothesis seven was, “There is no statistically significant relationship between the Autonomic Reactions subscale as measured by the Children’s Test Anxiety Scale (CTAS) and
students’ response time effort (RTE) while taking the Measures of Academic Progress® (MAP®) math assessment.” A Spearman correlation was used because the normality assumption for a Pearson correlation was violated. A very weak positive relationship was found between Autonomic Reactions and RTE ($r_s = .01, p = .86$). This finding failed to reject null hypothesis seven.

**Null Hypothesis Eight**

Null hypothesis eight was, “There is no statistically significant relationship between the Autonomic Reactions subscale as measured by the Children’s Test Anxiety Scale (CTAS) and students’ response time effort (RTE) while taking the Measures of Academic Progress® (MAP®) math assessment after controlling for gender.” A Spearman correlation was used because the normality assumption for a Pearson correlation was violated. A weak positive relationship was found in the female subsample between Autonomic Reactions and RTE for females ($r_s = .13, p = .19$). For the male subsample, a weak negative relationship was found between Autonomic Reactions and RTE ($r_s = -.09, p = .34$). These findings failed to reject null hypothesis eight.

Research question two asked, “Is there a relationship between any of the four measures of test anxiety as measured by the Children’s Test Anxiety Scale (CTAS) and the aggregate response time of students taking the Measures of Academic Progress® (MAP®) math assessment after separately controlling for gender?” There were a total of eight hypotheses related to this research question numbered nine through sixteen.

**Null Hypothesis Nine**

Null hypothesis nine was, “There is no statistically significant relationship between students’ Total Test Anxiety as measured by the Children’s Test Anxiety Scale (CTAS) and students’ aggregate response time (ART) while taking the Measures of Academic Progress®
(MAP®) math assessment.” A Spearman correlation was used because the normality assumption for a Pearson correlation was violated. A very weak positive relationship was found between Total Test Anxiety and ART ($r_s = .01, p = .88$). This finding failed to reject null hypothesis nine.

**Null Hypothesis Ten**

Null hypothesis ten was, “There is no statistically significant relationship between students’ Total Test Anxiety as measured by the Children’s Test Anxiety Scale (CTAS) and students’ aggregate response time (ART) while taking the Measures of Academic Progress® (MAP®) math assessment.” A Spearman correlation was used because the normality assumption for a Pearson correlation was violated. A very weak negative relationship was found in the female subsample between Total Test Anxiety and ART ($r_s = -.01, p = .95$). For the male subsample, a very weak positive relationship was found between Total Test Anxiety and ART ($r_s = .03, p = .77$). These findings failed to reject null hypothesis ten.

**Null Hypothesis Eleven**

Null hypothesis eleven was, “There is no statistically significant relationship between the Thoughts subscale as measured by the Children’s Test Anxiety Scale (CTAS) and students’ aggregate response time (ART) while taking the Measures of Academic Progress® (MAP®) math assessment.” A Spearman correlation was used because the normality assumption for a Pearson correlation was violated. A very weak positive relationship was found between Thoughts and ART ($r_s = .02, p = .81$). This finding failed to reject null hypothesis eleven.

**Null Hypothesis Twelve**

Null hypothesis twelve was, “There is no statistically significant relationship between the Thoughts subscale as measured by the Children’s Test Anxiety Scale (CTAS) and students’ aggregate response time (ART) while taking the Measures of Academic Progress® (MAP®) math
assessment after controlling for gender.” A Spearman correlation was used because the normality assumption for a Pearson correlation was violated. A very weak negative relationship was found in the female subsample between Thoughts and ART ($r_s = -.01, p = .92$). For the male subsample, a very weak positive relationship was found between Thoughts and ART ($r_s = .02, p = .83$). These findings failed to reject null hypothesis twelve.

**Null Hypothesis Thirteen**

Null hypothesis thirteen was, “There is no statistically significant relationship between the Off-Task Behaviors subscale as measured by the Children’s Test Anxiety Scale (CTAS) and students’ aggregate response time (ART) while taking the Measures of Academic Progress® (MAP®) math assessment.” A Spearman correlation was used because the normality assumption for a Pearson correlation was violated. A very weak negative relationship was found between Off-Task Behaviors and ART ($r_s = -.02, p = .76$). This finding failed to reject null hypothesis thirteen.

**Null Hypothesis Fourteen**

Null hypothesis fourteen was, “There is no statistically significant relationship between the Off-Task Behaviors subscale as measured by the Children’s Test Anxiety Scale (CTAS) and students’ aggregate response time (ART) while taking the Measures of Academic Progress® (MAP®) math assessment after controlling for gender.” A Spearman correlation was used because the normality assumption for a Pearson correlation was violated. A very weak negative relationship was found in the female subsample between Off-Task Behavior and ART ($r_s = -.03, p = .77$). For the male subsample, a very weak negative relationship was found between Off-Task Behavior and ART ($r_s = -.01, p = .94$). These findings failed to reject null hypothesis fourteen.
Null Hypothesis Fifteen

Null hypothesis fifteen was, “There is no statistically significant relationship between the Autonomic Reactions subscale as measured by the Children’s Test Anxiety Scale (CTAS) and students’ aggregate response time (ART) while taking the Measures of Academic Progress® (MAP®) math assessment.” A Spearman correlation was used because the normality assumption for a Pearson correlation was violated. A very weak positive relationship was found between Autonomic Reactions and ART ($r_s = .04, p = .53$). This finding failed to reject null hypothesis fifteen.

Null Hypothesis Sixteen

Null hypothesis sixteen was, “There is no statistically significant relationship between the Autonomic Reactions subscale as measured by the Children’s Test Anxiety Scale (CTAS) and students’ aggregate response time (ART) while taking the Measures of Academic Progress® (MAP®) math assessment after controlling for gender.” A Spearman correlation was used because the normality assumption for a Pearson correlation was violated. A very weak positive relationship was found in the female subsample between Autonomic Reactions and ART for females ($r_s = .02, p = .84$). For the male subsample, a very weak positive relationship was found between Autonomic Reactions and ART ($r_s = .05, p = .59$). These findings failed to reject null hypothesis sixteen.
CHAPTER FIVE: DISCUSSION, CONCLUSIONS, AND RECOMMENDATIONS

Discussion

The purpose of this nonexperimental correlational study was to determine the relationships between students’ response time, at the aggregate and also at the per item level (response time effort), while taking a computerized adaptive test, the Northwest Evaluation Association™ (NWEA™) Measures of Academic Progress® (MAP®) math test, and test anxiety, while controlling for gender for seventh grade math students at a school district in the West. This study addressed a gap in research regarding the relationship of test anxiety with aggregate response time and with response time effort (RTE) by utilizing a national standardized assessment and exploring and examining the relationships between response time and test anxiety. Most previous response time research utilized local assessments (Setzer et al., 2013). In addition, examinee effort (response time) was “one of the more subtle sources of error” (Setzer et al., 2013, p. 34) often disregarded as a confounding variable. In this chapter comparisons are made between the research literature and the findings of this study, conclusions and implications of the research are drawn, and a series of recommendations for future research are made.

Past test anxiety research evolved from researching test anxiety as a single dimension to more recent research studies that determined test anxiety to be multidimensional (Alpert & Haber, 1960; Hembree, 1988; Lowe et al., 2008; Mandler & Sarason, 1952). In addition, with advancements in technology, the response time of students on online assessments may be efficiently and objectively measured in a nonreactive and noninvasive manner. Two of the primary approaches to measure student effort on assessments have been self-report Likert scale measures and response time measures. Response time measures are seen as more objective
measures than self-report Likert scale measures and present an additional advantage in that they are able to ascertain student effort for both the overall test and on individual items (Wise et al., 2010).

Wise and Kong (2005) noted three distinct advantages of using computer-based tests regarding response time: a) the collection of the data is “unobtrusive and nonreactive” (p. 166), b) it is a “direction observation” (p. 166) of test-taker behavior, and c) response data is available for each individual item. One caveat of response time measures for individual items does depend on the nature of the assessment itself. Gathering response time data for individual items on a cognitive adaptive test is more complex and problematic than on a fixed form assessment since a cognitive adaptive test assigns significantly more unique items to a group of students as opposed to a fixed form assessment which may assign all test-takers the same 50 items, for example.

In using response time data from a cognitive adaptive test, this study applied a 10% normative threshold such that rapid-guessing behavior on an individual item if the time was at or below 10% of the mean time, rounded to the nearest 0.5 second, spent on all items by the participants in the study. The threshold for this study was set at 7.5 seconds as the participant mean time spent on each Northwest Evaluation Association™ (NWEA™) Measures of Academic Progress® (MAP®) test item was 76.5 seconds. The use of RTE must maintain a delicate balance between identifying an accurate number of rapid-guessing behaviors while simultaneously striving to avoid classifying solution behaviors as rapid-guessing behaviors. The former principle leads to a longer response time threshold, while the latter principle leads to a shorter response time threshold with the latter principle being the higher priority of the two principles (Wise et al., 2013).
Additional analyses in this study did not find a statistically significant difference in Total Test Anxiety between females and males, which contradicted the findings of previous test anxiety research findings which determined that females tended to have higher levels of test anxiety than males do (Cassady & Johnson, 2002; Datta, 2013; Fritts & Marszalek, 2010; Harpell & Andrews, 2013; Hembree, 1988; Lowe et al., 2011; Lowe & Lee, 2008; Segool et al., 2013). Furthermore, no statistically significant differences between females and males were found in the three Total Test Anxiety subscales of the Children’s Test Anxiety Scale (CTAS) survey: a) Thoughts, b) Off-Task Behaviors, and c) Autonomic Reactions. This finding was in contrast to previous test anxiety research which found that physiological arousal, which is related to Autonomic Reactions, was more prominent in females than in males, albeit with a small effect size (Lowe et al., 2011). Autonomic Reactions is closely related to physiological hyperarousal as both are categorized as forms of emotionality (Liebert & Morris, 1967).

Wise and DeMars (2010) found that males were much more likely than females to have RTE values below .90. Males were also found to have a statistically significantly lower average RTE than females, albeit with a small effect size (Setzer et al., 2013). Previous response time research by Wise et al. (2010) indicated that males demonstrated lower mean effort than females. While the mean effort of males in this study was lower than the mean effort of females, the mean difference was not statistically significant.

Given the nonnormality of the distribution of each of the variables used in the study, Spearman correlations were examined instead of Pearson correlations. Previous research supported the use of Pearson correlations with Likert-scale survey data (Jacqueline, 2013; Maurer & Pierce, 1998; Norman, 2010); however, with the presence of nonnormality in the data
in this study, Spearman correlations were calculated to determine if the Spearman results differed from the results obtained using Pearson correlations.

**Research Question One**

The first research question sought to examine the various relationships between each of the four measures of test anxiety, Total Test Anxiety, Thoughts, Off-Task Behaviors, and Autonomic Reactions, as measured by the Children’s Test Anxiety Scale (CTAS) and the response time effort (RTE) of students taking the Measures of Academic Progress® (MAP®) math assessment after separately controlling for gender. This led to the exploration of relationships for eight hypotheses, specifically hypotheses one through eight, in this study. Thus, each test anxiety measure was examined for a correlation with the RTE for the entire population sample and then also by gender.

The use of Pearson correlations was the initial approach chosen for this study; however, given the nonnormality of the distributions, Spearman correlations were calculated instead. The Spearman correlations found no statistically significant relationships among the eight hypotheses using response time effort (RTE) as the response time measure. The findings in this study failed to reject any of the first eight null hypotheses. Thus, when using Spearman correlations the response time measure (RTE) was not related to Total Test Anxiety, Thoughts, Off-Task Behaviors, or Autonomic Reactions in neither the total sample of participants nor by gender.

A weak negative relationship was found between Total Test Anxiety and RTE for the entire sample while no relationship was found between Total Test Anxiety and RTE for females, whereas a weak negative relationship was found between Total Test Anxiety and RTE for males. A weak negative relationship was also found between Thoughts and RTE for the entire sample, the females, and males. A very weak negative relationship was found between Off-Task
Behaviors and RTE for the entire sample, for females, and also for males. A very weak positive relationship was found between Autonomic Reactions and RTE for both the entire sample and for females; however, a weak negative relationship was found between Autonomic Reactions and RTE for males.

**Research Question Two**

The second research question sought to examine various relationships between each of the four measures of test anxiety, Total Test Anxiety, Thoughts, Off-Task Behaviors, and Autonomic Reactions, as measured by the Children’s Test Anxiety Scale (CTAS) and the aggregate response time of students taking the Measures of Academic Progress® (MAP®) math assessment after separately controlling for gender. This led to the exploration of relationships for eight hypotheses, specifically hypotheses nine through sixteen, in this study. Thus, each test anxiety measure was examined for a correlation with the aggregate response time for the entire population sample and then also by gender.

Although Pearson correlations were initially determined for each of the hypotheses nine through sixteen, again the nonnormality of the distributions led to the calculation of Spearman calculations. No statistically significant relationships were found between each of the four measures of test anxiety and the aggregate response times of students who took the MAP® math test. Furthermore, no such statistically significant relationships were found when controlling for gender.

A very weak positive relationship was found between Total Test Anxiety and ART for the entire sample and for females. A very weak positive relationship, however, was found between Total Test Anxiety and ART for males. A very weak positive relationship was found between Thoughts and ART for the entire sample and for males, while a very weak negative relationship
was found between Thoughts and ART for females. A very weak negative relationship was found between Off-Task Behaviors and ART for the entire sample, for females, and for males. A very weak positive relationship was found between Autonomic Reactions and ART for the entire sample, for females, and for males.

**Conclusions**

The initial approach of using Pearson correlations was abandoned in favor of calculating Spearman correlations due to normality assumption violations. While a vast majority of the relationships between the various test anxiety measures and RTE were negative and most of the relationships between the various test anxiety measures and ART were positive, none of the findings using Spearman correlations were statistically significant. In addition, all of the relationships had small effect sizes.

When comparing the means between males and females for Total Test Anxiety, Thoughts, Off-Task Behaviors, and Autonomic Reactions, it was found that females had higher means in all four test anxiety measures, which reflects the findings of Wren and Benson (2004), the authors of the Children’s Test Anxiety Scale (CTAS) survey used in this study. However, additional analyses found no statistically significant differences between males and females for Total Test Anxiety, Thoughts, Off-Task Behaviors, and Autonomic Reactions. Previous studies have found that females tend to have high levels of test anxiety than males (Cassady & Johnson, 2002; Datta, 2013; Fritts & Marszalek, 2010; Harpell & Andrews, 2013; Hembree, 1988; Lowe et al., 2011; Lowe & Lee, 2008; Segool et al., 2013).

The nonnormality and skewness of the response time data in this study concurs with findings by Setzer et al. (2013) who noted that the response time effort (RTE) score distribution was very negatively skewed. The findings by Setzer et al. (2013) indicated that a vast majority
of students gave sufficient effort given the large number of students’ with high RTE values.
Although nonnormality in this study was not entirely unanticipated, nonnormality does present a
conundrum when interpreting results that involve RTE values. Thus, this tempered the
significance of the statistical findings in this study. Nonetheless, the findings indicate that not
only does a potentially dynamic difference exist between RTE and ART but given the test
anxiety-RTE results revealed mostly negative relationship and test anxiety-ART relationships
revealed mostly positive relationships, the test anxiety-response time relationship warrants
further study.

Implications

There were a number of implications that resulted from the findings in this study, which
focused on helping close a gap in research regarding test anxiety-response time relationships.
While the results of this study did not reveal any significant findings, it is worth noting that RTE
presents a more dynamic way to measure student effort during the course of an assessment as test
anxiety relationships with ART were negligible in comparison to the test anxiety relationships
with RTE. Facilitating a testing environment with an emphasis on optimal effort per item should
be a focus of educators. This, of course, is predicated on the availability of technology assets
and the ability of educators to utilize assessments that are able to capture per item response
times.

Given the common presence of test anxiety among K-12 students, incorporating more test
anxiety interventions is a worthwhile endeavor. The test anxiety intervention treatments found to
be the most effective incorporated a blend of study skills treatment with another treatment, such
as a behavioral or a cognitive approach (Ergene, 2003; Hembree, 1988). Of the 21 different test
anxiety interventions examined in his meta-analysis, Ergene (2003) found the highest effect size
for a cognitive-study skills approach followed by a behavioral study-skills approach and then a behavioral approach, while the smallest effect sizes were found for meditation, exercise, and counseling approaches. Furthermore, the largest effect size for the time spent in test anxiety intervention therapy was found between 201 and 350 minutes. This was the optimal time spent in treatment, as the effect sizes declined for treatment times less than 201 minutes and greater than 350 minutes (Ergene, 2003). Thus, the type of treatment and the length of treatment factored into the effectiveness of a given test anxiety intervention. The integrated approach regarding test anxiety intervention strategies is a reflection of the complexity of the test anxiety construct itself.

An emphasis on test preparation, teaching time management to empower students to make efficient use of study time, and increasing student self-efficacy are other potential interventions for the average student (Bonaccio & Reeve, 2010). For example, informing students in advance about both the content on a test as well as the format of the testing items, such as multiple choice, constructed response, matching, and so forth, is a way to address students’ perceptions while also optimizing the test-taking experience (Zeidner, 1998).

To reiterate the difference between aggregate response time and response time effort (RTE), by definition, aggregate response time is the total time, or duration, a student spends to complete a test, whereas RTE is the proportion, represented as a decimal value, of test items for which a student exhibited solution behavior (Wise & Kong, 2005). One implication of this study was that RTE was a viable and dynamic factor in measuring student effort. One of the caveats of calculating RTE values is the arbitrary nature of normative thresholds themselves. This indicates a need for additional research in the area of RTE values and determining additional methods to measure RTE in a way that is as objective as possible.
Given the dynamic nature of response time effort (RTE), a greater emphasis should be placed on the student effort per item rather than on merely overall effort on an assessment. To gather accurate response time effort data, however, one must administer assessments online. The use of online assessments which provide per item response times for each test-taker would give practitioners additional insight into test-taker effort. In addition, the accuracy and validity of student growth measures is reliant on students putting forth their best effort on two separate testing occasions. The student growth measurement in and of itself is difficult to measure as it requires an assessment to be accurately measured at two different times and for all test-takers (Wise et al., 2013).

Wise et al. (2010) stated that the validity of test scores was dependent on two things: a) a well-constructed test and b) a test-taker who put forth his or her best effort. Low student effort “represents a serious potential threat to test score validity” (Wise & Kong, 2005, p. 163). Given the influence of test scores on district, school, educator, and student accountability and the use of test data to make informed instructional decisions, accounting for student effort as a confounding variable should be factored into test scores as “low examinee effort is a direct threat to test score validity” (Setzer et al., 2013, p. 35). Thus, given the dynamic nature of RTE data in this study and the noted impact of student effort on test score validity, a crucial implication is that practitioners strive to create a test-taking culture of maximum effort by all students for any and all assessments on each individual item.

**Limitations**

Limitations of this study included threats to internal and external validity. Instrumentation was a threat to internal validity since the CTAS taken by the participants was a self-report survey (Gall et al., 2007). Self-report measures themselves present limitations.
Participants were presumed to have provided authentic responses to each Likert-scale statement. It was also presumed that the students whose response times suggested inadequate effort by the participants on the NWEA™ MAp® math test were then sufficiently motivated to provide genuine responses to the items on the Children’s Test Anxiety Scale (CTAS) survey. Such surveys were found to be prone to response bias which has the potential to increase error variance (Merrell, 2008). In addition, the instrumentation of the NWEA™ math test itself was a threat since it is a cognitive adaptive test. Thus, no participant took the exact same test. The CTAS and NWEA™ results may also have increased variance due to fatigue.

Due to the somewhat arbitrary nature of how response time effort (RTE) was determined and also due to per item response times being rounded to the nearest second, the error variance increased and may have impacted the internal validity of the RTE values. The arbitrary nature of RTE calculations may have also contributed to the lack of data variance in the results of this study. Furthermore, universal RTE methodology has not yet been established. The methodologies used to determine item-level thresholds vary across studies, which complicates the comparison of results across RTE-related studies.

Experimental mortality was a concern due to attrition during this study as well (Gall et al., 2007). Not all students who agreed to participate did, in fact, participate. The test anxiety of students who left the study may have been different than the test anxiety of those students who participated. In an effort to maintain a high participation rate, test proctors provided a CTAS survey reminder to the study’s participants as they completed the NWEA™ math test. An additional limitation regarding test proctors is that the administrations of the CTAS survey and the NWEA™ math were conducted by multiple proctors rather than by one single proctor. The
may have led to inconsistencies in how the CTAS survey and the NWEA™ math tests were administered.

External threats to validity included population validity since non-Caucasian and low socioeconomic populations were underrepresented in the study compared to the non-Caucasian and low socioeconomic proportions represented in the general student K-12 population in the United States (Gall et al., 2007). Thus, results of this study may not be generalized to populations with different demographics, such as age, gender, grade level, ethnicity, socioeconomic status, and who reside in other geographic regions in the United States. The results of this study may also not be generalized to other content areas either.

**Recommendations for Future Research**

A number of recommendations are suggested for future replications or future variations of this study. Future studies of test anxiety and response time should include a larger sample size, a sample that is more ethnically and socioeconomically diverse, and include participants across multiple school sites and grade levels other than seventh grade. Wise et al. (2005) also found that different test proctors had a statistically significant impact on the effort of test-taking participants. The participants in this study were assigned to various teachers who proctored the test. Thus, this factor potentially had an impact on the aggregate response time and response time effort data used in this study. It would be ideal to replicate this study using one proctor to facilitate the assessment and test anxiety survey for all participants.

A common response time threshold was used for this study. As noted by (Setzer et al., 2013), this particular method does not necessarily capture threshold variations that result from the varying characteristics from different test items. Ideally, a unique threshold, Response Time Fidelity, is chosen for each individual item based on the response time distribution for that
particular item. That is more problematic for a cognitive adaptive assessment as opposed to a fixed form assessment. Wise and Kong (2005) found that the amount of reading on a test item influenced a student’s response time effort (RTE); controlling for the length of items would be necessary to make sure it was not also a confounding variable (Setzer et al., 2013).

The number of unique test items taken by all participants on a cognitive adaptive test could number in the thousands as opposed to a fixed form assessment where the number of unique test items taken by participants would be significantly less. Thus, replicating this study with a fixed form assessment rather than a cognitive adaptive assessment is suggested. Using a fixed form assessment would also lend itself well to conducting a response time study exploring the relationship between RTE and the Modified Caution Index and also the relationship between a measure of test anxiety and the Modified Caution Index.

It is recommended that future studies use a different normative threshold (NT), such as NT15 or NT20, which would set thresholds at 15% or 20%, respectively, of the mean time per item. It is far more practical to obtain data from fixed form assessments than cognitive adaptive assessments to determine normative thresholds. The caveat with using higher NT thresholds is that it becomes more likely that non-effortful responses are included for responses that are, in fact, effortful. Setzer et al. (2013) recommended the use of bimodal distributions with an RTE cap of 10 seconds as the NT measure. This would be a worthwhile NT measure to utilize for a future test anxiety-RTE study. Wise & Kong (2005) also noted that multiple choice correct answers should also be balanced as test-takers tended to choose the middle options (answers b and c) more often in an effort to eliminate that as a confounding variable when determining RTE.

Although Wise et al. (2010) found that the mean effort of students did not vary by the time of the year and did not vary by the day of the week, the mean effort of students decreased
throughout the day. Thus the time of day could be removed as a confounding variable by assessing students at the same time of day since the time of day was shown to influence student effort with effort experiencing a more significant decline in the latter part of the day (Wise et al., 2005). The mean effort decreased across grade levels, and males demonstrated lower mean effort than females with the mean effort difference between males and females increasing as the day progressed. More recent research, however, noted that RTE values tended to be higher in the spring than in the fall and suggested that rapid-guessing behavior was more prevalent in the fall than in the spring (Wise et al., 2013); thus, replicating this study in the spring would be an additional recommendation. Furthermore, the study could be done using the NWEA™ MAP® Reading test in lieu of the Math test. This would, of course, need to be conducted with assistance by and the support of NWEA™, who so graciously provided response time data for this study.

The Test Anxiety Inventory for Children and Adolescents (TAICA), when it becomes commercially available, could be used in lieu of the Children’s Test Anxiety Scale (CTAS) survey to replicate this study between test anxiety and RTE. It measures seven dimensions, including overall test anxiety with four test anxiety subscales, a Lie scale, and a Facilitating measure of test anxiety, which measures test anxiety that enhances student achievement (Lowe & Lee, 2008). Given that self-report measures do tend to have error variance, biofeedback measures of test anxiety, such as heart rate and blood pressure, would be an alternate manner of determining test anxiety that participants experience throughout an assessment. This data could be gathered simultaneously with response time data to explore potential relationships between test anxiety and response time effort in real-time.
Very few test anxiety intervention studies have been conducted in the United States and abroad, especially in K-12 education (Ergene, 2003; von der Embse et al., 2013). Von der Embse, Barterian and Segool (2013) indicated that as few as 10 such viable K-12 test anxiety intervention studies were conducted between 2000 and 2010. Of those 10 research studies, only four were conducted in the United States. The studies examined Cognitive Behavioral Therapy (CBT), relaxation techniques, biofeedback, and mixed methods approaches. Research has shown that, while various interventions have been discussed, there also is a tremendous need to conduct studies designed to better identify students with elevated levels of debilitating test anxiety and to also specifically examine debilitating test anxiety interventions (Ergene, 2003; von der Embse et al., 2013). One caveat of intervention strategies is that a one-size-fits-all approach is highly unlikely to be effective for all student populations (von der Embse et al., 2013). One caveat of intervention strategies is that a one-size-fits-all approach is highly unlikely to be effective for all student populations (von der Embse et al., 2013). This reinforces the multimodal approach touted by Zeidner (1998). For example, the differences in how females and males react to anxiety and situational cues suggests that controlling for gender when studying test anxiety interventions may be necessary (Reeve & Bonaccio, 2008).

In conclusion, a number of small correlations were found between RTE and various test anxiety measures while no correlations were found between aggregate response time and test anxiety measures. The development of and use of more multidimensional test anxiety surveys will enhance the understanding and treatment of test anxiety. In addition, while test anxiety measures currently focus on summed values, identifying cut points to differentiate between what constitutes high, medium, and low test anxiety levels would be a worthwhile endeavor. Additional RTE research is also needed, with an emphasis on the potential development of a
universal RTE method. This would not only reduce error variance but also strengthen the quality of the interpretation of results in students involving RTE.
REFERENCES


Ackerman, P. L., & Heggestad, E. D. (1997). Intelligence, personality, and interests: Evidence for overlapping traits. *Psychological Bulletin, 121*(2), 219-245. doi:10.1037/0033-2909.121.2.219


Lawrence Earlbaum Associates.

at the meeting of the American Educational Research Association, New York, NY.

and Behavioural Difficulties, 8(2), 101-107. doi:10.1080/13632750300507010

Covington, M. V. & Omelich, C. L. (1987). “I knew it was cold before the exam”: A test of the 
anxiety-blockage hypothesis. Journal of Educational Psychology, 79, 393- 400. 
doi:10.1037/0022-0663.79.4.393


relationships between test anxiety and related variables. International Journal of 

Datta, P. (2013). Is test anxiety a peril for students with intellectual disabilities? Journal of 

Dickhäuser, O., Reinhard, M., & Englert, C. (2011). “Of course I will . . . ”: The combined effect 
of certainty and level of expectancies on persistence and performance. Social Psychology 


doi:10.1080/10615800701847823


doi:10.1348/000709906X161704


doi:10.1080/01443410701452264


Reeve, C. L., Heggestad, E. D., & Lievens, F. (2009). Modeling the impact of test anxiety and test familiarity on the criterion-related validity of cognitive ability tests. *Intelligence, 37*(1), 34-41. doi:10.1016/j.intell.2008.05.003


responses to high-stakes testing. *Psychology in the Schools, 50*(5), 489-499.

doi:10.1002/pits.21689


Wilhelm, O., & Schulze, R. (2002). The relation of speeded and unspeeded reasoning with mental speed. *Intelligence, 30*(6), 537. doi:10.1016/S0160-2896(02)00086-7


APPENDICES

Appendix A: Children’s Test Anxiety Scale

(Wren & Benson, 2004)

This survey will be administered with Survey Monkey and will be titled Test Attitude Survey.

Circle the answers that best describe or tell about you.

1. Enter your district-issued student identification number __________

**SAMPLE** – Please read the following statement and decide if it describes the way you are while you are taking tests. If the statement is **almost never** or **never** like you, you should circle (select) 1. If the statement describes the way you are **some of the time**, circle (mark) 2. If the statement describes the way you are **most of the time**, circle (mark) 3. If the statement is **almost always** or **always** like you, circle (mark) 4.

<table>
<thead>
<tr>
<th>While I am taking tests ...</th>
<th>ALMOST NEVER</th>
<th>SOME OF THE TIME</th>
<th>MOST OF THE TIME</th>
<th>ALMOST ALWAYS</th>
</tr>
</thead>
<tbody>
<tr>
<td>I think about doing other things.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
</tbody>
</table>

The rest of the items describe how some students may think, feel, or act while they are taking tests. Please, read each statement carefully and circle (select) the answer that best describes the way you are while you are taking a test. Remember that there are no “right” or “wrong” answers on this survey. Please give truthful answers.

<table>
<thead>
<tr>
<th>While I am taking tests ...</th>
<th>ALMOST NEVER</th>
<th>SOME OF THE TIME</th>
<th>MOST OF THE TIME</th>
<th>ALMOST ALWAYS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. I wonder if I will pass</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>2. My heart beats fast</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>3. I look around the room</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>4. I feel nervous</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>5. I think I am going to get a bad score</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>While I am taking tests ...</td>
<td>ALMOST NEVER</td>
<td>SOME OF THE TIME</td>
<td>MOST OF THE TIME</td>
<td>ALMOST ALWAYS</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>--------------</td>
<td>-----------------</td>
<td>-----------------</td>
<td>--------------</td>
</tr>
<tr>
<td>6. It is hard for me to remember the answers</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>7. I play with my pencil</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>8. My face feels hot</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>9. I worry about failing</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>10. My belly feels funny</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>11. I worry about doing something wrong</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>12. I check the time</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>13. I think about what my score will be</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>14. I find it hard to sit still</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>15. I wonder if my answers are right</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>16. I think I should have studied more</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>17. My head hurts</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>18. I look at other people</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>19. I think most of my answers are wrong</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>20. I feel warm</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>While I am taking tests ...</td>
<td>NEVER</td>
<td>THE TIME</td>
<td>THE TIME</td>
<td>ALWAYS</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>-------</td>
<td>----------</td>
<td>----------</td>
<td>--------</td>
</tr>
<tr>
<td>21. I worry about how hard the test is</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>22. I try to finish up fast</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>23. My hand shakes</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>24. I think about what will happen if I fail</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>25. I have to go to the bathroom</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>While I am taking tests ...</th>
<th>ALMOST NEVER</th>
<th>SOME OF THE TIME</th>
<th>MOST OF THE TIME</th>
<th>ALMOST ALWAYS</th>
</tr>
</thead>
<tbody>
<tr>
<td>26. I tap my feet</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>27. I think about how poorly I am doing</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>28. I feel scared</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>29. I worry about what my parents will say</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>30. I stare</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
</tbody>
</table>

Wren & Benson (2004). Children’s Test Anxiety Scale. © Used and adapted with permission from Dr. Douglas G. Wren, Ed.D.
## TEST ATTITUDE SURVEY

Circle the answers that best describe or tell about you.

1. I am a... boy.  girl.
2. Circle your grade.  3rd  4th  5th  6th

**SAMPLE** – Please read the following statement and decide if it describes the way you are while you are taking tests. If the statement is *almost never* or *never* like you, you should circle 1. If the statement describes the way you are *some of the time*, circle 2. If the statement describes the way you are *most of the time*, circle 3. If the statement is *almost always* or *always* like you, circle 4.

<table>
<thead>
<tr>
<th>While I am taking tests...</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>I think about doing other things.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The rest of the items describe how some students may think, feel, or act while they are taking tests. Please read each statement carefully and decide if the statement describes how you think, feel, or act during a test. Then circle the answer that best describes the way you are while you are taking a test. If you are not sure which answer to circle, read the statement again before circling your answer. Remember that there are no “right” or “wrong” answers on this survey. Please give truthful answers.
## While I am taking tests...

<table>
<thead>
<tr>
<th></th>
<th>ALMOST NEVER</th>
<th>SOME OF THE TIME</th>
<th>MOST OF THE TIME</th>
<th>ALMOST ALWAYS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. I wonder if I will pass.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>2. My heart beats fast.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>3. I look around the room.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>4. I feel nervous.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>5. I think I am going to get a bad grade.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
</tbody>
</table>

## While I am taking tests...

<table>
<thead>
<tr>
<th></th>
<th>ALMOST NEVER</th>
<th>SOME OF THE TIME</th>
<th>MOST OF THE TIME</th>
<th>ALMOST ALWAYS</th>
</tr>
</thead>
<tbody>
<tr>
<td>6. It is hard for me to remember the answers.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>7. I play with my pencil.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>8. My face feels hot.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>9. I worry about failing.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>10. My belly feels funny.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
</tbody>
</table>

## While I am taking tests...

<table>
<thead>
<tr>
<th></th>
<th>ALMOST NEVER</th>
<th>SOME OF THE TIME</th>
<th>MOST OF THE TIME</th>
<th>ALMOST ALWAYS</th>
</tr>
</thead>
<tbody>
<tr>
<td>11. I worry about doing something wrong.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>12. I check the time.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>13. I think about what my grade will be.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>14. I find it hard to sit still.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>15. I wonder if my answers are right.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Statement</td>
<td>Almost Never</td>
<td>Some of the Time</td>
<td>Most of the Time</td>
<td>Almost Always</td>
</tr>
<tr>
<td>------------------------------------------------</td>
<td>--------------</td>
<td>------------------</td>
<td>------------------</td>
<td>---------------</td>
</tr>
<tr>
<td>16. I think that I should have studied more.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>17. My head hurts.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>18. I look at other people.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>19. I think most of my answers are wrong.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>20. I feel warm.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>21. I worry about how hard the test is.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>22. I try to finish up fast.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>23. My hand shakes.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>24. I think about what will happen if I fail.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>25. I have to go to the bathroom.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>26. I tap my feet.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>27. I think about how poorly I am doing.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>28. I feel scared.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>29. I worry about what my parents will say.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>30. I stare.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
</tbody>
</table>

**Thank you for your help!**
Appendix C: Permission Letter to Reproduce and Publish the Children’s Test Anxiety Scale

Brom, Michael

From: Douglas G. Wren <Douglas.Wren@VBSchools.com>
Sent: Saturday, July 18, 2015 8:43 AM
To: Brom, Michael
Subject: Re: CTAS Survey

Hi Michael,

Congratulations on receiving approval from your IRB.

I reviewed the adapted version of the Children's Test Anxiety Scale (CTAS) and am in full agreement that your adaptation of the CTAS is appropriate and acceptable. In addition, you have my permission to reproduce and publish the original CTAS as well as your adapted version in the appendices of your dissertation.

Feel free to contact me if you need any further information. Good luck as you proceed with your study.

Doug Wren

Douglas G. Wren, Ed.D.
Assessment Specialist
Department of Teaching & Learning
Virginia Beach City Public Schools

From: Brom, Michael <mbrom@liberty.edu>
Sent: Thursday, July 16, 2015 4:38 PM
To: Douglas G. Wren
Subject: RE: CTAS Survey

Good afternoon, Dr. Wren,

I have received IRB approval for my test anxiety study to proceed.

My participants will be taking the cognitive adaptive online test, NWEA’s Measures of Academic Progress math test, as the assessment in my study.

The wording of some of the CTAS questions say “grade” – NWEA assigns scores. To avoid confusion for participants, I have replaced “grade” with “score” and have attached the adapted version (as it would appear in my dissertation) for you to examine. I want to ensure that it is an acceptable adaptation of the survey before proceeding.

In addition, the publishing department at my school, Liberty University, requires that doctoral students receive both permission to use and permission to reproduce/publish the survey used in a study (in the Appendix section of the dissertation so readers know exactly what survey was used).

Please, let me know if the wording adaptation meets with your approval and if I may include the CTAS survey (I can include the original survey for one Appendix document and the adapted survey for another Appendix document).
Appendix D: Institutional Review Board Approval Letter

Brom, Michael

From: IRB, IRB <IRB@liberty.edu>
Sent: Tuesday, June 9, 2015 11:42 AM
To: Brom, Michael
Cc: Hutter, Sarah C (School of Education); Garzon, Fernando (Ctr for Counseling & Family Studies); IRB, IRB
Subject: IRB Approval 2228.060915: A Correlational Analysis of Test Anxiety and Response Time on a Computerized Adaptive Math Test among Seventh-Grade Students by Gender
Attachments: BromApproval_06_15.pdf; Brom_2228StampedAssent.pdf; Brom_2228StampedConsent.pdf; Annual Review Form.docx; ChangeInProtocol.docx

Dear Michael,

We are pleased to inform you that your study has been approved by the Liberty IRB. This approval is extended to you for one year from the date provided above with your protocol number. If data collection proceeds past one year, or if you make changes in the methodology as it pertains to human subjects, you must submit an appropriate update form to the IRB. The forms for these cases are attached to your approval email.

Your IRB-approved, stamped consent form is also attached. This form should be copied and used to gain the consent of your research participants. If you plan to provide your consent information electronically, the contents of the attached consent document should be made available without alteration.

Please retain this letter for your records. Also, if you are conducting research as part of the requirements for a master’s thesis or doctoral dissertation, this approval letter should be included as an appendix to your completed thesis or dissertation.

Thank you for your cooperation with the IRB, and we wish you well with your research project.

Sincerely,

Fernando Garzon, Psy.D.
Professor, IRB Chair
Counseling

(434) 592-4054

Liberty University | Training Champions for Christ since 1972
Appendix E: Participant Consent Form

CONSENT FORM

A Correlational Analysis of Test Anxiety and Time on a Computer Adaptive Math Test among Seventh-Grade Students by Gender
Michael Brom, Principal Investigator
Liberty University
School of Education

Your child is invited to be in a research study of test anxiety and the amount of time students spend taking the Northwest Evaluation Association Measures of Academic Progress (MAP) math test. Your child was selected as a possible participant because he or she is a seventh-grade student in the Lewis-Palmer School District. I ask that you read this form and ask any questions you may have before agreeing to permit your child to participate in the study.

Michael Brom, a doctoral candidate in the School of Education at Liberty University, is conducting this study.

Background Information:
The purpose of this study is to examine the relationships between test anxiety and the response time of students taking a computer adaptive math test. The study will also examine the test anxiety and response time relationships by gender. The researcher wishes to determine the relationships that exist between the levels of test anxiety students experience on a computer adaptive math test and the amount of time students spend taking the test. The amount of time will be examined in two different ways: a) the total amount of time a student spends taking the test and b) the amount of time a student spends on each item on the test (response time effort).

Procedures:
If you agree to allow your child to be in this study, I would ask him/her to do the following things:
- Take the Children’s Test Anxiety Scale (CTAS) survey via Survey Monkey immediately following completion of the NWEA MAP math test. It will take participants approximately 5-10 minutes to complete this survey.

Risks and Benefits of being in the Study:
The study has risks: a) students who participate in the study may feel more anxious than usual knowing that test anxiety is a factor being examined in the study and b) students may be embarrassed by poor performance knowing that their data, while still confidential, is being used in the study. There is a possibility that participants will experience one or both of the aforementioned risks. There are no direct benefits to your child for participating in this study and the risks are minimal.

Compensation:
Because student motivation is a variable of interest as measured by aggregate response time and response time effort in the study, no incentive or compensation will be provided.
to participants in order to not influence student effort, which could affect the findings in the study.

**Confidentiality:**
The records of this study will be kept private and will be password protected. In any sort of report I might publish, I will not include any information that will make it possible to identify a participant in the study. The Northwest Evaluation Association (NWEA) Measures of Academic Progress (MAP) math test data will be stored by NWEA and Lewis-Palmer School District. The Children’s Test Anxiety Survey (CTAS) data will be stored online and accessible by the Lewis-Palmer School District Assessment office and the researcher and will be password-protected. Once the MAP math test and the CTAS data have been merged by the Lewis-Palmer Assessment office by student ID numbers, the student IDs will be replaced with randomly generated identification numbers to protect the privacy of each participant in the study. The research records will be stored securely and only the researcher will have access to the records. The data will be password-protected and stored on the researcher’s laptop hard drive for a period of no more than three years.

**Voluntary Nature of the Study:**
Participation in this study is voluntary. Your decision whether or not to allow your child to participate will not affect his or her current or future relations with Liberty University, Lewis-Palmer Middle School, or the Lewis-Palmer School District. If you decide to allow him/her to participate, he/she will be free to not answer any question or withdraw at any time without affecting those relationships.

**Contacts and Questions:**
The researcher conducting this study is Michael Brom. You may ask any questions you have now. If you have questions later, you are encouraged to contact him at mbrom@liberty.edu, (720) 470-7969, or his dissertation chairperson, Dr. Sarah Hutter, at schutter@liberty.edu.

If you have any questions or concerns regarding this study and would like to talk to someone other than the researcher, you are encouraged to contact the Institutional Review Board, 1971 University Blvd, Suite 1837, Lynchburg, VA 24515 or email at irb@liberty.edu.

**Please notify the researcher if you would like a copy of this information to keep for your records.**

**Statement of Consent:**
I have read and understood the above information. I have asked questions and have received answers. I consent to participate in the study.
The Liberty University Institutional Review Board has approved this document for use from 6/9/15 to 6/8/16.
Protocol # 2226.060915

(NOTE: DO NOT AGREE TO PARTICIPATE UNLESS IRB APPROVAL INFORMATION WITH CURRENT DATES HAS BEEN ADDED TO THIS DOCUMENT.)

Signature: ____________________________ Date: ____________

Signature of parent or guardian: ________________ Date: ____________
(If minors are involved)

Signature of Investigator: __________________________ Date: ____________
Appendix F: Assent Form

Review Board has approved
this document for use from
6/9/15 to 6/8/16
Protocol # 22238 060915

Assent of Child to Participate in a Research Study

What is the name of the study and who is doing the study?
A Correlational Analysis of Test Anxiety and Time on a Computer Adaptive Math Test among Seventh-Grade Students by Gender. The principal investigator for the study is Mr. Michael Brom.

Why are we doing this study?
We want to study whether or not a student’s test anxiety (fear or worry) makes a difference in how much time students spend taking a test.

Why are we asking you to be in this study?
You are being asked to be in this research study because we want to know more about the relationship between test anxiety and how much time middle school students spend taking online tests.

If you agree, what will happen?
If you are in this study, you will take a Test Attitude Survey once you have completed the NWEA MAP math test. The Test Attitude Survey will take 5-10 minutes to complete.

Do you have to be in this study?
No, you do not have to be in this study. If you want to be in this study, then tell the researcher. If you don’t want to, it’s OK to say no. The researcher will not be angry. You can say yes now and change your mind later. It’s up to you.

Do you have any questions?
You can ask questions any time. You can ask now. You can ask later. You can talk to the researcher. If you do not understand something, please ask the researcher to explain it to you again.

Signing your name below means that you want to be in the study.

Signature of Child  Date

Michael Brom, mbrom@liberty.edu
Dr. Sarah Hutter, schutter@liberty.edu
Liberty University Institutional Review Board,
1971 University Blvd, Suite 1837, Lynchburg, VA 24515
or email at jrb@liberty.edu
Appendix G: Recruitment Letter

Dear Parent/Guardian:

I want to welcome you and your child to Lewis-Palmer Middle School for the 2015-16 school year. I am entering my third year as a math teacher at LPMS and my 24th year overall in education. For the past two years I have been working on my dissertation, with a keen interest in test anxiety. I am conducting research to better understand test anxiety and student motivation. The purpose of my research is to examine relationships between test anxiety and student response time on the Northwest Evaluation Association (NWEA) Measures of Academic Progress (MAP) math test. This study was submitted to district personnel for consideration and approved. I am emailing you to invite your child to participate in this local study.

On Monday August 24th, all 7th grade students at LPMS will take the NWEA MAP math test. This assessment is one that is used for local purposes. The only additional requirement for the study I am conducting is that students complete a test anxiety survey, the Children's Test Anxiety Scale (CTAS), immediately following the NWEA MAP math test. It will take students 5-10 minutes to complete the test anxiety survey online at SurveyMonkey.com. Your child's district-issued student identification number will be needed as part of his or her participation on the test anxiety survey. This data will only be used to merge test anxiety data with NWEA MAP math test response time data, will be done by district personnel, and will remain completely confidential. Once the data have been merged, the district-issued student identification numbers will be replaced by randomly generated numbers to ensure the confidentiality and anonymity for all participants. The data to be analyzed will not contain any student names whatsoever.

A consent form (page 3) and assent form (page 4) are found in the attachment to this email. In order for your child to participate in this study, pages 3 and 4 must be signed, completed, and returned. You may scan and email the signed forms to me or have your child return the signed forms to their math teacher by Friday, August 21.

A primary local benefit of this study is that it will indicate the overall test anxiety level our 7th grade students experience, which will help us teachers improve upon the instructional and assessment strategies we use in our classrooms on a regular basis. If you have questions or concerns, I will be available all day during 7th grade processing at LPMS on Tuesday, August 11, and will hold an informational meeting on Thursday, August 13, at 7 pm at LPMS. You may also reach me by email at mbrom@lewispalmer.org, or you may call my personal cell phone at [Phone Number].

Thank you for your consideration,
Michael Brom
Lewis-Palmer Middle School Math Teacher, 7 Red
Doctor of Education Student, Liberty University
Appendix H: Institutional Review Board Change in Protocol Letter

Brom, Michael

From: IRB, IRB
Sent: Thursday, August 6, 2015 11:30 AM
To: Brom, Michael
Cc: Hutter, Sarah C (School of Education); Garzon, Fernando (Ctr for Counseling & Family Studies); IRB, IRB
Subject: IRB Change in Protocol Approval: IRB Approval 2228.060915: A Correlational Analysis of Test Anxiety and Response Time on a Computerized Adaptive Math Test among Seventh-Grade Students by Gender

Good Afternoon Michael,

This email is to inform you that your request to “add a recruitment letter to email to parents, as well as making [yourself] available to parents/students during . . . 7th grade processing day on August 11 to answer questions/concerns about the study and to obtain signatures for the consent/assent forms” and to “attach the IRB approved consent/assent forms to [your] email to parents” has been approved. Thank you for submitting your recruitment email for our review and documentation.

Thank you for complying with the IRB’s requirements for making changes to your approved study. Please do not hesitate to contact us with any questions.

We wish you well as you continue with your research.

Best,

G. Michele Baker, MA, CIP
Administrative Chair of Institutional Research
The Graduate School
(434) 592-5530

LIBERTY UNIVERSITY
Liberty University | Training Champions for Christ since 1971