This is to certify that the
dissertation entitled

TEST ANXIETY ASSOCIATED WITH HIGH-STAKES TESTING AMONG ELEMENTARY SCHOOL CHILDREN: PREVALENCE, PREDICTORS, AND RELATIONSHIP TO STUDENT PERFORMANCE

presented by

NATASHA KATHERINE SEGOOL

has been accepted towards fulfillment
of the requirements for the

Ph.D. degree in School Psychology

Major Professor's Signature

6/3/09

Date

TEST ANXIETY ASSOCIATED WITH HIGH-STAKES TESTING AMONG ELEMENTARY SCHOOL CHILDREN: PREVALENCE, PREDICTORS, AND RELATIONSHIP TO STUDENT PERFORMANCE

By

Natasha Katherine Segool

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

School Psychology

2009

ABSTRACT

TEST ANXIETY ASSOCIATED WITH HIGH-STAKES TESTING AMONG ELEMENTARY SCHOOL CHILDREN: PREVALENCE, PREDICTORS, AND RELATIONSHIP TO STUDENT PERFORMANCE

By

Natasha Katherine Segool

The current study explored differences in test anxiety on high-stakes standardized achievement testing and classroom testing among elementary school children. This is the first study to directly examine differences in student test anxiety across two testing conditions with different stakes among young children. Three hundred and thirty-five students in grades three through five participated in the current study. Wilcoxon Signed Rank Tests revealed that as a whole, students reported significantly more overall test anxiety in relation to high-stakes testing than classroom testing on two different measures of test anxiety (r = -.21 and r = -.10). Students also reported significantly more cognitive (r = -.20) and physiological (r = -.24) symptoms of test anxiety in relation to high-stakes testing than classroom testing. Teachers' perceptions of student anxiety were also examined in the current study and consistent with student self-reported test anxiety, results indicated that teachers believed students experienced significantly more anticipatory anxiety in relation to high-stakes testing versus classroom testing.

Analyses of variance indicated that low test-anxious students performed significantly better on high-stakes testing while there was no difference in performance between moderate and high test-anxious students. Further, multiple regression analyses indicated that test anxiety contributed significantly to English Language Arts test performance among third and fourth grade students and Math test performance among

third grade students. Additionally, the current study examined the relationship between test anxiety and student demographic characteristics. Results of multiple regression analyses indicated that student gender and grade significantly predicted student test anxiety, while student ethnicity, educational verification, and socioeconomic status did not.

The current study also examined the relationship between three measures of test anxiety and differences in how three test anxiety classification systems sorted students into low, moderate, and high test-anxious groups. Correlation analyses indicated that the Children's Test Anxiety Scale for Children (CTAS), Test Anxiety Scale for Children (TASC), and Behavior Assessment Scale for Children, Second Edition, Test Anxiety subscale (BASC-2-TA) were strongly related to one another ($rs > .70$). However, chi-square goodness of fit tests revealed that there were significant differences in the proportions of students classified as low, moderate, and high test-anxious across classification systems. Using the CTAS and TASC to classify students, prevalence rates of high test anxiety in relation to classroom testing ranged from 11 to 21 percent, respectively. Similar rates of students were classified as highly test-anxious across the high-stakes and classroom testing conditions using the CTAS.

ACKNOWLEDGEMENTS

I would like to acknowledge several individuals who provided much appreciated support to me throughout my graduate schooling. First, my thanks are extended to Dr. John Carlson, my advisor and dissertation chair, who encouraged and supported me to pursue my passion of applied psychosocial research throughout my studies at Michigan State University. Additionally, I want to thank my dissertation committee members, Drs. Sara Bolt, John Kosciulek, and Ed Roeber, for their time, thoughtful critiques, and interest in this project. My appreciation is expressed to the school teachers, staff, and administrators who provided invaluable support in completing this research project. Additionally, I want to thank all of the school children and families who participated in this project. My heartfelt gratitude also extends to Anisa Goforth who diligently coordinated the data collection effort and to Justin Barterian who contributed to data collection and entry efforts. Finally, I would like to thank my family for their unwavering interest, support, and care throughout my graduate studies. This research project was partially funded by the Michigan Association of School Psychologists Research Grant and the School Psychology Program Student Research Award. Additional support was provided by the Leadership Training Grant Fellowship from the U.S. Department of Education, Office of Special Education Programs.

ACKNOWLEDGEMENTS

I would like to acknowledge several individuals who provided much appreciated support to me throughout my graduate schooling. First, my thanks are extended to Dr. John Carlson, my advisor and dissertation chair, who encouraged and supported me to pursue my passion of applied psychosocial research throughout my studies at Michigan State University. Additionally, I want to thank my dissertation committee members, Drs. Sara Bolt, John Kosciulek, and Ed Roeber, for their time, thoughtful critiques, and interest in this project. My appreciation is expressed to the school teachers, staff, and administrators who provided invaluable support in completing this research project. Additionally, I want to thank all of the school children and families who participated in this project. My heartfelt gratitude also extends to Anisa Goforth who diligently coordinated the data collection effort and to Justin Barterian who contributed to data collection and entry efforts. Finally, I would like to thank my family for their unwavering interest, support, and care throughout my graduate studies. This research project was partially funded by the Michigan Association of School Psychologists Research Grant and the School Psychology Program Student Research Award. Additional support was provided by the Leadership Training Grant Fellowship from the U.S. Department of Education, Office of Special Education Programs.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Test anxiety is associated with significant impairment in performance on academic achievement assessments, lower courses grades, decreased motivation, and increased stress (Cizek & Berg, 2006; Hembree, 1988, Zeidner, 1998). Test anxiety is comprised of psychological, physiological, and behavioral reactions that occur in association with concern about the negative outcomes resulting from failure or poor performance in evaluative situations (Zeidner). In comparison to low test-anxious students, students with high levels of test anxiety score lower on tests of academic achievement in core academic areas including reading, English, and mathematics (Hembree). Additionally, students' grade point averages (GPA) are inversely related to test anxiety levels, and significantly more test anxious students drop out of school than non-test anxious students (Cizek & Berg; Hembree; Spielberger, 1966).

The prevalence rate of impairing levels of test anxiety among young children has been estimated to range from as little as 10% to as much 40% of school-aged children (King & Ollendick, 1989; Turner, Beidel, Hughes, & Turner, 1993). Female students and non-Caucasian students consistently report higher prevalence rates of test anxiety (Hembree, 1988; Putwain, 2007; Turner et al., 1993). Higher prevalence estimates of test anxiety in contemporary society have been attributed to the increased testing that children are exposed to in public schools. As Sarason foreshadowed in 1959, "we live in a test-conscious, test-giving culture in which the lives of people are in part determined by their test performance" (p.26).

The impact that testing has on the educational and occupational outcomes of

children in the United States has increased as a result of the educational accountability movement, which has dramatically increased the prevalence of standardized testing for public school children in elementary, middle, and high school. The No Child Left Behind Act of 2001 (NCLB) requires annual testing and reporting of statewide academic achievement assessments in the areas of reading and mathematics during grades three through eight and once in high school. Testing in the content area of science is done once in elementary, middle, and high school. These testing requirements mean that students take no fewer than seventeen state-mandated standardized achievement assessments prior to high school graduation (U.S. Congress, 2002). NCLB testing has high-stakes for educators and communities because student achievement data are publicly reported and directly linked to school funding, administration, and employment decisions. However, students' psychological reactions to and perceptions of this testing is less clear.

Although there is a clear need to understand the relationship between student test anxiety and testing required by NCLB, relatively little research has been conducted that directly examines the relationship between test anxiety and federally mandated annual achievement assessments. A number of researchers have examined the perceptions that teachers, parents, and school administrators have about the impact of large-scale testing programs on students; however, most of these studies are indirect assessments that may have been influenced by the respondents' beliefs, concerns, or worries about high-stakes assessments. Indirect perception studies suggest that state testing programs have resulted in increased student anxiety, increased stress, lowered motivation (Abrams, Pedulla, & Madaus, 2003; Barksdale-Ladd & Thomas, 2000; Jones & Egley, 2004; 2006; Jones et al., 1999), increased focus on test preparation (Abrams et al.; Barksdale-Ladd & Thomas;

Jones et al.), and increased job stress and lowered job satisfaction for teachers (Barksdale-Ladd & Thomas; Jones et. al; Jones & Egley, 2004; 2006). While these studies provide important data about how these stakeholders perceive the impact of statewide testing programs, there is a need to study the impact of testing programs on children directly in order to understand how students are actually impacted by testing programs.

The few existing studies that directly examine how students are impacted by high-stakes testing are equivocal and limited by methodological issues. One projective study of students' feeling about testing found that students overwhelmingly felt stress, anxiety, worry, and isolation as a result of testing (Triplett & Barksdale, 2005). Alternatively, Mulvenon and colleagues (2001; 2005) found that most students felt positively about testing and that test anxiety, when present, did not impair student performance on the state assessment. However, these findings must be interpreted with caution because the measure of test anxiety used in these studies was not empirically validated and did not reflect current conceptual models of test anxiety. Most recently, Putwain (2008) found that although secondary students in the United Kingdom reported higher levels of test anxiety in relation to low- versus medium- and high-stakes examinations, there was a small but significant negative relationship between test anxiety and performance on all examinations. These contradictory findings across studies highlight the need to systematically examine the relationship between test anxiety and high-stakes testing among students using empirically validated measures of test anxiety and methodologically sound data collection techniques.

This dissertation study examined how young children perceive and experience

high-stakes standardized achievement testing. By surveying elementary school students to assess test anxiety in relation to both classroom testing and state achievement testing required by NCLB, the levels and prevalence rates of low, moderate, and high test anxiety associated with high-stakes testing in comparison to classroom testing were explored. Also, teachers' perceptions of students' responses to the different testing conditions were examined. Next, the relationship between test anxiety and test performance was examined. Additionally, student demographic characteristics associated with test anxiety were examined. Finally, the relationship between three measures of childhood test anxiety was examined. This study significantly adds to and expands upon the current literature on test anxiety by directly examining the relationship between test anxiety and high-stakes testing among young students who are required by law to complete annual standardized achievement assessments.

CHAPTER 2

LITERATURE REVIEW

High levels of test anxiety are associated with detrimental outcomes, including impaired learning and performance on recall tasks, achievement measures, and intellectual assessments (Hembree, 1988; Zeidner, 1998). The current political and educational climate in the United States has resulted in the frequent use of standardized academic achievement assessments to measure student performance and educational quality at the school, community, and state level. This increased emphasis on testing necessitates a more complete understanding of the impact that test anxiety has on students' performance in high-stakes contexts. This literature review will provide a context for this study by a) defining test anxiety, its presentation, and its effects on students, b) examining the educational accountability movement and the increasing emphasis of testing in American public schools, and c) critiquing the existing literature on test anxiety and assessments administered for educational accountability. Finally, the rationale for the current dissertation study will be presented along with the current study's research questions and hypotheses.

Test Anxiety

Test anxiety is defined as the psychological, physiological, and behavioral responses associated with concern about experiencing negative outcomes as a result of failure or poor performance on an assessment or in an evaluative situation (Cizek & Berg, 2006; Zeidner, 1998). In explaining the need to study test anxiety, early test anxiety researcher Seymour Sarason (1959) argued that "we live in a test-conscious, test-giving culture in which the lives of people are in part determined by their test performance"

(p.26). Currently, educational testing programs have an even more significant impact on the lives of children, and a subset of children respond to testing programs with high levels of test anxiety. In turn, children's performance on these tests may be negatively affected by test anxiety (Meijer, 2001). Therefore, there is a pressing need to consider how students perceive testing programs and how their testing performance is affected by test anxiety. The following sections will review the construct of test anxiety, the prevalence rates and developmental patterns of test anxiety, and the relationship between test anxiety and learning and academic performance.

Theoretical Models

Yerkes and Dodson (1908) published a now classic work on the facilitating and debilitating effects of arousal on learning. They found that test performance was enhanced when low levels of arousal were present, whereas high levels of arousal resulted in performance decrements. The tipping point between facilitating and debilitating levels of arousal shifted depending on the task's difficulty, with impairment resulting at lower arousal levels on more difficult tasks. This work anticipated the study of test anxiety, which focuses on the impact of anxious arousal in testing situations.

In the 1950s, Seymour Sarason formalized the study of test anxiety and developed standardized measures to examine its cognitive and affective components. Sarason and his colleagues hypothesized that high test-anxious individuals experience increased worry and autonomic arousal during testing situations. Further, they suggested that this worry and arousal results in impaired task performance (Zeidner, 1998). Liebert and Morris (1967) advanced Sarason's theoretical model of test anxiety by examining how cognitive and physiological aspects of test anxiety differentially affect test performance. Their

model defined worry as the cognitive concerns related to fear of failure and emotionality as the affective or autonomic responses to testing situations (Stöber & Pekrun, 2004). Liebert and Morris found that worry was more strongly related to test performance than emotionality, suggesting that the impairment associated with test anxiety results from cognitive interference rather than physiological arousal.

Later, Spielberger's distinction between state and trait anxiety led to further investigation about the specific relationship between testing performance and each anxiety type (Stöber & Pekrun, 2004). State anxiety is transitory and results from a specific situation whereas trait anxiety is a relatively stable personality trait. Spielberger and colleagues (1976) found that underlying trait anxiety predisposes individuals to situation-specific test anxiety when faced with a performance task. Spielberger's work suggests that trait anxiety scales that measure underlying anxiousness and worry may be more closely related to test performance than general state anxiety scales that focus more on querying physiological arousal (Zeidner, 1998). Later research by Irwin Sarason confirmed that cognitive aspects of anxiety such as self-critical beliefs, high self-focus, and a focus on the competitive nature of assessments are more impairing to test performance than autonomic arousal (Zeidner). The cognitive model of test anxiety suggests that excessive focus on the self and the task result in impaired attention, memory, and retrieval during testing, which ultimately results in impaired test performance (Smith, Ingram, & Brehm, 1983; Wine, 1971). For example, Matters and Burnett (2003) found that task-irrelevant thinking is associated with a greater rate of omitting answers to items on a high-stakes achievement test, suggesting that the cognitive aspects of test anxiety impede students' testing behaviors and strategies.

More recently, Zeidner (1997) proposed an integrative transactional model of test anxiety that considers the impact of personal, task, and situational factors on test anxiety, rather than focusing entirely on person-specific factors, such as worry and emotionality. The transactional model considers how task-specific factors, such as the test difficulty, the testing atmosphere, the examiner characteristics, the presence of time constraints, and the mode of administration, and person-specific factors, such as trait test anxiety, self-efficacy, scholastic ability, previous testing experience, and need for achievement, all influence the responses that individuals have to testing situations (Zeidner). The transactional model emphasizes the importance of cognition in the development and expression of test anxiety by suggesting that person- and task-specific factors affect the cognitive appraisal an individual makes about the test situation (Zeidner). In turn, this appraisal influences the level of test anxiety experienced by the individual.

Lowe and colleagues (2008) have expanded Zeidner's transactional model of test anxiety using a biopsychosocial framework. This model encompasses Zeidner's transactional model while also recognizing that societal factors, such as schools, communities, parents, and families also influence the development and expression of test anxiety. This biopsychosocial model is particularly relevant given the increasingly important role that educational testing programs have in evaluating schools and educators. The model highlights three different processes involved in the expression of test anxiety, including the individual's behavior, cognition, and physiology. Behaviors include both task-relevant and task-irrelevant behavior (e.g., focusing attention on task or skimming through items). Physiological reactions include emotional arousal (e.g., increased heart rate, rapid breathing, or muscle tightness). Cognitions include worry that

interferes with the task (e.g., thoughts about social humiliation or the consequences of failure). Each of these behavioral, cognitive, and physiological responses affect task performance, and in turn, affect how the individual responds to future testing experiences (Lowe et al.). Lowe and colleagues' biopsychosocial model of test anxiety and Zeidner's transactional model of test anxiety have been combined and adapted to serve as the conceptual framework for the present study (see Figure 1).

Prevalence Rates

Estimates on the prevalence rate of test anxiety among school-aged children have varied widely (Zeidner, 1998). Traditional measures of test anxiety have not included norms with specific cut-points for diagnosing clinical levels of test anxiety. Therefore, researchers have used different criteria to define clinical levels of test anxiety, which may explain the range in test anxiety prevalence estimates. King and Ollendick (1989) report that the prevalence of test anxiety among school-aged children may range from as little as 10% to as much as 30%. However, they estimate that the rate of children who experience clinically significant impairment is likely to be on the lower end of this spectrum (King & Ollendick). Similarly, Hill and Wigfield (1984) suggest that between two and three children in typical classrooms, or approximately ten percent of children, are highly test-anxious and experience impairments in test performance as a result. More recently, some researchers have used traditional scales and developed their own operational definitions of high test anxiety in order to measure the prevalence rate of test anxiety among children.

For example, Turner and colleagues (1993) conducted an analysis of test anxiety among eight to twelve year-old African American school children. Children were

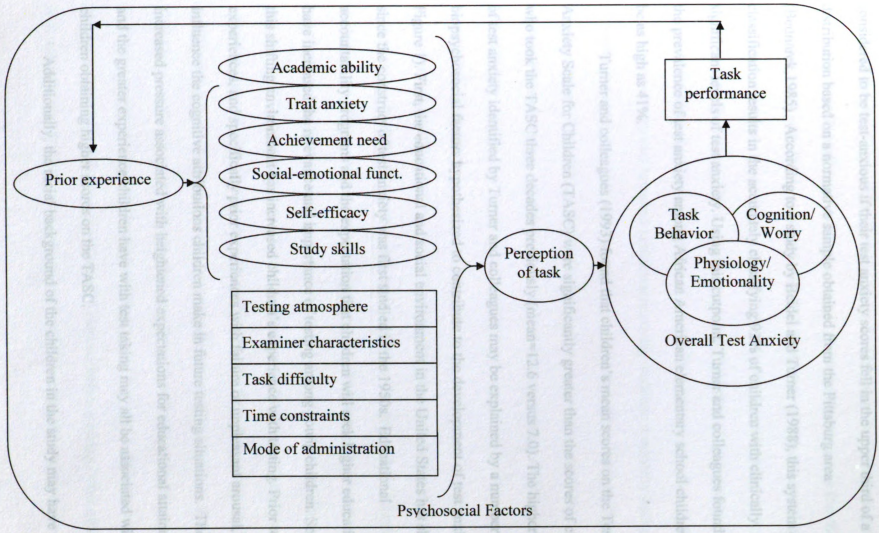


Figure 1: Conceptual Framework adapted from Zeidner, 1997 and Lowe et al., 2008

considered to be test-anxious if their test anxiety scores fell in the upper third of a distribution based on a normative sample obtained from the Pittsburg area (Bednarek,1985). According to a study by Beidel and Turner (1988), this system of classification results in the accurately classifying 92% of children with clinically significant levels of test anxiety. Using this cutpoint, Turner and colleagues found that the prevalence of test anxiety among African American elementary school children may be as high as 41%.

Turner and colleagues (1993) found that children's mean scores on the Test Anxiety Scale for Children (TASC) were significantly greater than the scores of children who took the TASC three decades previously (mean=12.6 versus 7.0). The higher levels of test anxiety identified by Turner and colleagues may be explained by a number of biopsychosocial factors hypothesized to contribute to the development of test anxiety (see Figure 1). First, the educational and social environment in the United States has shifted since the construct of test anxiety was first studied in the 1950s. Educational accountability programs and the expectation that children will seek higher educations have increased the relevance and importance of testing among young children. Second, this shifting environment has increased children's experience with testing. Prior testing experiences, and specifically prior experiences with failure or unpleasant arousal, influence the cognitive attributions children make in future testing situations. The increased pressure associated with heightened expectations for educational attainment and the greater experience children have with test taking may all be associated with children obtaining higher scores on the TASC.

Additionally, the racial background of the children in the study may have

influenced the rate of test anxiety found in the study. Stereotype threat is the fear that one's behavior will confirm an existing stereotype about the identity group that the person identifies with. Steele and Aronson (1995) found that when race was salient to intellectual test-takers, the performance of African Americans was impaired in comparison to Caucasians. This finding suggests that if children of color perceive educational testing situations as a measure of worth with respect to race, the result may be increased anxiety and impaired performance among students of minority backgrounds. However, the extent to which stereotype threat is related to performance-based anxiety remains unclear (Ryan & Ryan, 2005). Further studies are needed to clarify the relationship between race, stereotype threat, and test anxiety in young children in the United States.

Demographic Patterns

Putwain (2007) recently examined demographic patterns in the expression of test anxiety among 1348 14-16 year-old students in the United Kingdom. Putwain examined students during their last two years of compulsory schooling because they take examinations that greatly impact their educational and occupational futures during this time. Students' gender, socioeconomic status, ethnicity, and language of origin were all related to the level of test anxiety experienced by students. First, females reported significantly greater overall test anxiety, worry, and emotionality than males (Putwain). This finding replicates previous studies that have found females consistently report higher test anxiety than males. Hembree (1988) reported that gender differences in test anxiety emerge around first grade and steadily increase before peaking in the middle and high school years. Differences in test anxiety between females and males were moderate

in grades 5-10 (Effect size; ES = 0.43) and small in grades 1 and 2 (ES = 0.14), grades 3 and 4 (ES = 0.28), and grades 11, 12, and postsecondary (ES = 0.27; Hembree). The question remains whether this difference may be attributed to differences in socialization processes that result in females more openly disclosing anxious symptoms or whether female students actually experience greater levels of anxiety (Wigfield & Eccles, 1989).

Second, Black, Asian, and ethnic minority students reported significantly greater overall test anxiety, worry, and emotionality than Caucasian students (Putwain, 2007). This finding is supported by the findings of Turner and colleagues (1993) and Wren and Benson (2004) who found higher rates of test anxiety among African American elementary school children, including more autonomic reactions and worry-related thoughts. Alternatively, Beidel and colleagues (1994) found that rates of test anxiety did not differ significantly by racial group, although African American children with test anxiety had greater rates of social phobia diagnoses than Caucasian children with test anxiety. In interpreting these different findings, is important to recognize that the cultural and experiential differences between the samples from the United States and the United Kingdom might affect the relationship between ethnicity and test anxiety (Bodas & Ollendick, 2005). Similarly, the different ages of the children examined across these studies highlights the need for further examination of the relationship between ethnicity and test anxiety across all school-aged children. Hembree (1988) found an interaction between age and ethnicity in the expression of test anxiety, with significantly greater rates of test anxiety among Black children in grades two to four (ES = 0.52) and grades five to eight (ES = 0.21) than Whites, but no differences between older Black and White high school students. Alternatively, Hispanic students in grades 4-12 reported

significantly greater test anxiety than White students (ES = 0.36; Hembree).

Finally, Putwain (2007) examined differences in test anxiety in relation to high-stakes testing among students from different socioeconomic backgrounds (SES), including those with a head of household employed in routine/manual, intermediate, managerial/professional, or not classified jobs such as retirement or full-time study. Students from routine/manual socioeconomic backgrounds reported significantly higher overall test anxiety, worry, and emotionality than students from managerial/professional backgrounds. Although students from intermediate and not classified backgrounds did not report significantly different overall test anxiety, both reported significantly greater worry than students from managerial/professional backgrounds (Putwain). Putwain's findings are consistent with Hembree (1988)'s meta-analytic finding that there is a small but significant relationship between SES and test anxiety (ES = -.13). These findings suggest that the socioeconomic status of students impacts the expression of test anxiety, with particular differences found between low and high socioeconomic groups. However, differences in cognitive responses to testing also exist between middle and upper socioeconomic groups.

Other researchers have examined the relationship between age and test anxiety. Rates of test anxiety appear to increase over the course of elementary school and become more stable over the middle and high school years (Hembree, 1988, Hill & Sarason, 1966; McDonald, 2001; Sarason et al., 1960). According to a meta-analysis of test anxiety differences between students in adjacent grades from 2 to 12, test anxiety becomes established by second grade and rates increase steadily until fifth grade before remaining relatively constant through the end of high school and then declining during

college (Hembree). However, the reported decline in test anxiety during the college years may be the result of increased attrition among students with greater test anxiety (Spielberger, 1966).

Increases in test anxiety over the school-age years may be related to children's increased exposure to testing over time and to the conditioning that occurs following repeated concern about test failure, pressure to perform well, or feedback about poor performance (King & Ollendick, 1989). Given the ever-increasing focus on using testing performance to make high-stakes decisions, it is reasonable to hypothesize that the overall prevalence and absolute level of test-anxiety will continue to increase among young students over time (Hill & Wigfield, 1984). Students are now exposed to more frequent assessments, assessments begin at a younger age, and assessments are directly linked to rewards and sanctions. Thus, in this age of educational accountability, prevalence rates and demographic patterns of test anxiety among young children need to be reexamined.

Effects on Performance and Educational Outcomes

The detrimental outcomes associated with test anxiety are extensive, including impaired cognitive functioning, scholastic underperformance, and limited educational and occupational outcomes (Zeidner, 1998). In comparison to low test-anxious individuals, high test-anxious individuals experience significant impairment in evaluative testing situations that arouse worry. In laboratory settings, a strong inverse relationship exists between task performance and the amount of time spent worrying during the task (Holroyd, Westbrook, Wolf, & Badhorn, 1978). This relationship has also been found in applied settings. For example, in a meta-analysis on 1112 high school and college

students, there was a significant correlation between worry and impaired performance on aptitude and achievement assessments (r = -0.31) and on course grades (r = -0.26; Hembree, 1988). Emotionality was also correlated to a lesser degree with significant levels of impairment on aptitude and achievement assessments (r = -0.15) and on course grades (r = -0.19; Hembree). These findings reflect the relationship that exists between cognitive and physiological aspects of test anxiety and test performance. Similar findings by other researchers suggest high test-anxious students would be expected to score significantly lower than low test-anxious students on real-world achievement tests (Seipp, 1991; Zeidner).

Over time, the relationship between test anxiety and test performance translates into impaired educational and occupational outcomes for test-anxious students. A meta-analysis of over 6000 students across all grade levels found that high versus low test anxiety scores were related to significantly lower scores on intelligence (IQ), achievement, and aptitude measures (ES = -0.48; Hembree, 1988). Although lower IQ scores could be suggestive of lower ability in students, it is important to note that intelligence assessments require testing, thus the lower scores may be more reflective of impairment from test anxiety than lower innate ability among these students. For example, Spielberger (1966) found that although students' grade point averages (GPA) in a single semester did not differ significantly, suggesting that high and low anxious students had commensurate academic ability, more than 20% of the high-anxious college students eventually dropped out of college as a result of academic failure, whereas fewer than six percent of low anxious students experienced the same outcome. Spielberger's findings on students falling in the upper and lower 20% of the distribution on a measure

of manifest anxiety, suggest that there is a cumulative negative effect of anxiety on academic performance. These studies are important because they suggest that test anxiety is associated with multiple types of impairment in educational functioning and attainment. Unfortunately, the majority of test anxiety research has been conducted on college students, and relatively less is known about test anxiety in young children. This pattern of research limits our understanding about the short- and long- term impact that test anxiety has on young children.

Seymour Sarason led early research efforts in the field of test anxiety, and his work on test anxiety among elementary school children has remained highly influential. In 1964, Sarason, Hill, and Zimbardo conducted a seminal, longitudinal study on the relationship between test anxiety and student performance on intellectual and achievement tests. Sarason and colleagues (1964) reported on 713 students who were followed from first and second grade until fourth and fifth grade, respectively. Their data convincingly showed that test anxiety had a significant detrimental impact on students' acquisition of knowledge and students' utilization of knowledge on evaluative tasks (Sarason et al.). Additionally, they found that children's level of test anxiety increased with age and that high test-anxious children experienced less growth over time in academic achievement than low test-anxious children (Sarason et al.). For example, when children were categorized into high and low test anxiety groups on the basis of their third grade test anxiety scores, the high test-anxious group's reading achievement scores were eight to fourteen months behind the low test-anxious group's scores.

When Sarason, Hill, and Zimbardo (1964) examined the relationship between students' verbal intelligence using the Lorge-Thorndike Intelligence Scale, achievement

measures, and test anxiety, they found that the impairment high test-anxious children showed on achievement measures was not fully explained away by differences in intelligence. This finding strongly suggests that test anxiety impairs students' initial learning and/or utilization of skills, resulting in impaired academic performance. Similarly, in a meta-analysis of early childhood studies, Hembree (1988) found that second grade achievement in reading and English was significantly and inversely related to test anxiety (r = -0.15), although there was no significant relationship between math and test anxiety in grade two. In a much larger sample of older students from grade three to postsecondary school, both reading and mathematics achievement was significantly and inversely related to test anxiety (r = -0.24 and r = -0.22, respectively). These findings highlight the need for careful examination of how test anxiety affects students' skill development and skill utilization during evaluative situations. Additional research that considers student ability while also examining the impact of test anxiety on performance is needed.

In their follow-up report, Hill and Sarason (1966) analyzed five years of longitudinal data on 670 students in order to examine test anxiety stability and the relationship between test anxiety and school achievement. While there was low stability in test anxiety status from first to fifth grade (rs = 0.15 and 0.20 for boys and girls, respectively), there was significantly greater stability in test anxiety status by late-elementary school from fourth to sixth grade, (rs = 0.58 and 0.49 for boys and girls, respectively; Hill & Sarason). Hill and Sarason also found that the absolute level of test anxiety significantly increased among children from first to fifth grade. Similarly, Hembree (1988) found that there was a significant relationship between test anxiety level

and grade level, with test anxiety levels increasing as students advanced in elementary school from grade to grade (ES = -0.24). This pattern of increasing test anxiety in both level and stability over the elementary school years suggests that the impairment associated with test anxiety may similarly increase over time.

Further, Hill and Sarason (1966) found that the strength of the negative correlation between test anxiety and achievement increased over the elementary school years. Students' academic achievement was significantly more impaired by test anxiety as students advanced in elementary school. Even when high and low test-anxious children were matched on the basis of intelligence scores, low test-anxious boys performed significantly better on measures of overall achievement, reading, arithmetic-comprehension, and arithmetic-problem-solving, and low test-anxious girls scored significantly better on measures of reading achievement (Hill & Sarason). Children's test anxiety scores were also significantly related to their grades, with low test-anxious children receiving better grades than high test-anxious children (Hill & Sarason). The data showing that intelligence does not explain away differences between high and low test-anxious children's academic achievement strongly suggests that students with high test anxiety experience significant impairment in their educational development over the course of their education as a result of test anxiety. These data highlight the need for further study of test anxiety among young children as well as examination of early intervention techniques that may intercede in the developmental trajectory of test anxiety in young children.

Educational Accountability Movement

In light of the established relationship between test anxiety and impaired

academic achievement, it is important to consider how the educational accountability movement in the United States has resulted in an increased use and focus on test performance in schools. The ongoing national focus on educational reform and accountability developed in response to the 1983 publication of *A Nation at Risk*. In response to concerns that students in the United States were lagging behind international competitors in mathematics and science, the National Commission on Excellence in Education (NCEE) provided a scathing condemnation of the then current educational practices. The report found that the educational system in the United States was lacking essential quality features and that students' performance lagged behind other industrialized countries in core content areas (NCEE, 1983). The NCEE concluded the report with a series of recommendations for educational reform that were designed to enhance student learning and performance. *A Nation at Risk* stimulated legislative action that fundamentally shifted the role that the federal government took in overseeing education in the United States. It marked the beginning of a national focus on standards-based reform in education by calling for rigorous and measurable achievement standards that could be monitored and evaluated.

Historically, the development, administration, implementation, and evaluation of educational services have been duties performed by state and local educational agencies. The federal government's role in education was focused on ensuring that states complied with federal procedural rules and regulations designed to provide students with appropriate access to educational services. Following *A Nation at Risk*, however, the federal government took an increasingly active role in developing performance-based accountability policies to ensure that educational reform measures were implemented

(Wong & Nicotera, 2007). For example, in 1989, President George H. Bush and the National Governor's Association met for a summit on education and began to develop educational performance standards. The work accomplished at this summit eventually became the basis for the Goals 2000: Educate America Act, which delineated national education goals for what students should know and be able to do (U.S. Congress, 1994a). These goals reflected a shift in national educational philosophy from one that dictated minimum competency standards to one that expected all students to meet high educational standards when provided with appropriate opportunities to learn. Further, Goals 2000 mandated national assessment practices to evaluate student achievement in comparison to national and state performance goals.

Additionally, the federal government increasingly mandated performance-based educational reform measures through successive reauthorizations of the Elementary and Secondary Education Act (ESEA). The 1994 reauthorization of ESEA known as the Improving America's Schools Act (IASA) focused on developing stronger accountability measures for student achievement, including a specific focus on at-risk students through a redesign of Title I (U.S. Congress, 1994b). IASA mandated the use of statewide assessment programs to measure students' annual yearly progress (AYP). Measuring students' AYP was designed to enhance academic and curriculum standards in schools by focusing local and national attention on school performance. Most recently, when ESEA was reauthorized in January 2002, the influence and importance of statewide academic achievement testing programs were further increased by attaching financial and other high-stakes sanctions and rewards to schools' performance on the assessments.

No Child Left Behind Act of 2001

The No Child Left Behind Act of 2001 (NCLB), or the 2002 reauthorization of ESEA, mandated the statewide administration of annual academic performance assessments in order to hold public schools accountable for the education of school children and to "close the achievement gap with accountability, flexibility, and choice, so that no child is left behind" (US. Congress, 2002, p. 1). In order to achieve this goal, NCLB sought to enhance educational services by promoting alignment between state academic achievement standards and state academic assessments that are used to make decisions about school funding, employment, and other major administrative functions. The state assessment programs required by NCLB are specifically designed for "identifying and turning around low-performing schools that have failed to provide a high-quality education to their students" (US. Congress, 2002, p. 16). While NCLB did not establish a national curriculum for schools, all states were mandated to develop educational accountability plans. Additionally, states were obligated to develop a single assessment system that measures the annual yearly progress of all students on the state's academic and curricular standards. States must "include sanctions and rewards, such as bonuses and recognition" to hold local schools accountable for student achievement in the areas of reading or language arts, mathematics, and science (US. Congress, 2002, p. 16). The requirement that test scores be publicly reported and linked to rewards and sanctions such as school funding and employment decisions makes this testing high-stakes in nature for educators and communities. Although individual students' scores are not linked directly to sanctions and rewards, the high stakes associated with this testing for educators may result in teachers and schools placing increased pressure on students to

perform well on the test.

Reading and mathematics assessments must be administered annually to students in grades 3-8 and once during grades 10-12. Science assessments must be administered at least once during grades 3-5, grades 6-9, and grades 10-12. In addition, states are free to assess other content areas such as social studies. Hence, NCLB requires students to take no fewer than 17 assessments that are directly linked to high-stakes decisions over the course of their elementary and secondary education. When students' scores are reported on these frequent assessments, scores are disaggregated into groups on the basis of students' major social characteristics, including major racial and ethnic groups, economic status, disability status, and limited English proficiency status. Decisions about schools are made by examining student achievement data in total and in disaggregate. By disaggregating the scores, the stakes for students who traditionally have been underserved by the educational system are effectively increased since schools may receive sanctions for failing to meet annual yearly progress standards for any of the disaggregated groups.

Additionally, summary reports of disaggregated assessment data must be publicly reported, and itemized scores are reported to parents, teachers, principals, and administrators. Although major educational organizations including the National Association of School Psychologists (NASP) and the American Educational Research Association (AERA) warn against using standardized assessments for *individual* decision-making, the NCLB requirement that scores be reported individually for each student implicitly encourages educators and parents to use the data to make individual academic decisions about students (AERA, 2000; NASP, 2003). This reporting requirement increases the testing stakes for students, who may be evaluated by parents

and teachers on the basis of their state assessment performance.

Michigan Educational Assessment Program

In the state of Michigan, the Michigan Educational Assessment Program (MEAP) has been adapted to comply with the requirements of NCLB. The MEAP was initially developed in 1969 to determine what students across the state knew and could do in relation to the educational standards set by the state. The MEAP was traditionally administered at key points in students' academic careers. As a measure of students' skills in the core content areas of mathematics, reading, science, social studies, and writing, the MEAP was designed to be an accountability measure for use in school improvement by indicating curriculum and instructional needs at schools. The MEAP was specifically designed to be interpreted in relation to other achievement measures and instructional variables (State of Michigan, 2004a).

In its current form, the MEAP is a criterion-referenced test that is administered during the fall of each academic year. Testing occurs over a period of two to three weeks and it requires approximately eight hours to complete the assessment in English Language Arts, Mathematics, and Science. MEAP scores are used as the key indicator of student achievement in Michigan. In order to meet the Adequate Yearly Progress (AYP) requirements of NCLB, student groups, schools, and districts are evaluated on their MEAP performance in relation to annual performance goals. Scores are reported at the individual student level through reports sent to parents and maintained in student files, at the classroom level through class roster reports, at the school level through summary reports disaggregated by student groups, and at the district and state levels. Student performance in each content area is classified as advanced, proficient, partially proficient,

or not proficient, and shifts in performance levels from year to year are also monitored and reported (Griffiths & Peterman, 2007).

Additionally, school and district level scores are posted on the State's Department of Education website and are also publicly reported on through the news media (e.g., newspaper, television, and internet). The public reporting of MEAP scores at the school and district levels has important implications for communities, including, but not limited to influencing real estate values and decisions families make about where to live. Similarly, in some cases, the individual student performance data that is reported to schools and parents directly influences educational decisions made on behalf of children, such as grade retention and special education referral decisions, (e.g., Fielding, 2004).

Although the Michigan Department of Education cautions against using the MEAP as a sole measure of educational achievement, in accordance with the federal No Child Left Behind Act, high-stakes outcomes are linked directly to MEAP scores in Michigan. Sanctions associated with not meeting AYP standards increase with each successive year of failure. For example, following two successive years of failure, Title I schools must notify parents of their status, must offer school choice and pay for transportation costs totaling up to twenty percent of Title I funds, must develop school improvement plans, and must use ten percent of Title I funds for professional development. Following three years of failure, schools must offer students supplementary educational services in addition to the aforementioned sanctions. Following four years of failure, schools must take corrective action such as staff replacement, school restructuring, or new school management. Thus, the sanctions associated with not meeting AYP significantly affect students, educators, and

communities; and the sanctions increase in magnitude with each successive year of failure (State of Michigan, 2004b).

In sum, the educational accountability movement has resulted in an increased emphasis on using performance-based measures to evaluate and make decisions about students and schools. The MEAP has a significant impact on students, families, schools, and communities. In terms of immediate impacts, students and schools must devote a significant amount of time, energy, and funding to completing the MEAP. Following administration of the MEAP, high-stakes consequences are associated with students' performance on the MEAP. These consequences exist at individual, family, school, and community levels, and by influencing decisions such as individual students' educational placements, family decisions about what communities to live in and send children to school in, and school decisions about teacher employment, administrative structuring, and funding allotments.

Test Anxiety and High-Stakes Assessment

In 1984, Hill and Wigfield predicted that the trend towards using test scores to evaluate educational programs and to make decisions about individual students would result in increasing rates of test anxiety among students. The authors argued,

> test scores have a strong influence in determining children's school progress and what they will be able to do later in life.... Because of testing's importance, we believe it is essential that all student's test scores accurately reflect what they do know and not be negatively influenced by test anxiety and other test-taking factors (Hill & Wigfield, 1984, p. 122).

The federal mandate to use standardized, high-stakes testing programs to evaluate schools

and students rests on the assumption that students' test scores accurately reflect students' knowledge and skills. However, as Hill and Wigfield (1984) suggest, this assumption must be demonstrated empirically since other extraneous factors, including test anxiety, may differentially affect students' performance on these assessments.

In light of the evidence that test anxiety is associated with impaired learning and use of knowledge in evaluative situations, it follows that test anxiety may similarly impact student performance on the standardized tests of academic achievement required by NCLB (Cizek, 2001; Meijer, 2001). The following sections critically review the existing literature on the impact of high-stakes assessments on children. Studies are organized by methodology, with the few studies directly examining students' perceptions of assessments reviewed first and indirect studies examining the perceptions that teachers, parents, and other school personnel have about the impact of testing on students presented second. Gaps in the literature are also examined in the context of presenting a rationale for further study of test anxiety and its relationship to academic achievement assessments mandated by NCLB.

Direct Studies on the Impact of Testing on Children

Only four studies have directly examined children's perceptions of achievement assessments mandated by NCLB. These studies offer important insights into how students experience these testing programs. Triplett and Barksdale (2005) examined 225 third through sixth graders' drawings and written responses to probes that asked them to first draw pictures about their testing experience and then describe their pictures through writing. The authors concluded that the children overwhelmingly experienced negative emotions, including nervousness and anger, in relation to testing. Students also expressed

feelings of isolation from their peers and teachers. In contrast to their hypothesis that the high-stakes assessment experience would be "basically meaningless" (p. 257), testing was a pervasively negative experience for students and the students exhibited an awareness of and concern about the high-stakes associated with the assessment. The authors suggest that the negative valance of the children's creative products might stem from test anxiety; however, the study did not directly measure test anxiety and the study methodology required the researchers to interpret the meaning of children's drawings and writing samples, introducing the possibility of inaccurate interpretations. Therefore, while this study provides interesting initial data suggesting that children experience significant stress during testing, it needs to be followed up with a more direct examination of the relationship between high-stakes assessment and test anxiety.

In contrast to the findings of Triplett and Barksdale, two studies conducted by Mulvenon and colleagues (2001; 2005) suggest that the prevalence and impact of test anxiety among students taking high-stakes assessments is minimal. In their initial study, Mulvenon, Connors, and Lenares (2001) surveyed 283 fifth grade students who had just completed state mandated Stanford Achievement Test- Ninth Edition (SAT-9) testing about their attitudes towards testing. Seventy-five percent of students reported having positive attitudes toward testing, and they did not exhibit stress or impaired performance. Although there was a significant relationship between reported pressure and lower scores on the SAT-9, pressure accounted for less than two percent of the variance in SAT-9 scores, suggesting that pressure was not a clinically significant predictor of test performance.

In a later report, Mulvenon, Stegman, and Ritter (2005) expanded on their initial

report and examined the perceptions that teachers, principals, counselors, students, and parents had about the relationship between high-stakes testing and student performance. The authors reported that the pressure students perceive from parents and teachers, and not test anxiety, was significantly related to poorer test performance. Additionally, the authors found that teachers' anxiety level was not significantly related to student performance, which contrasts concerns raised by Triplett and Barksdale (2005) about the negative impact of teachers' anxious actions and statements. These studies suggest that concern about test anxiety in relation to high-stakes testing may be unnecessary.

However, the studies conducted by Mulvenon and colleagues (2001; 2005) had a number of methodological weaknesses that call their findings into question. First, the studies used a 24-item investigator-developed survey to measure six constructs, including test anxiety, rewards, pressure, attitude toward testing, and self-efficacy in reading and math. While there were eleven questions about test anxiety, this subscale's concurrent validity was not established with established measures of test anxiety, and all of the other constructs were measured using limited items ranging from one to three questions. Additionally, when examining the relationship between test anxiety and test performance, the authors controlled for student ability, which is important for isolating the effect of test anxiety on test performance. However, the authors controlled for student ability using students' scores on a criterion-referenced examination from the previous spring. Since the SAT-9 and the spring exam were both standardized, group-administered assessments, students may have also perceived the spring assessment to be anxiety-provoking, which would invalidate the use of this assessment as a measure of ability unaffected by test anxiety. These studies findings need to be replicated using an empirically validated

measure of test anxiety and a more representative measure of student ability.

Most recently, Putwain (2008) examined whether the stakes of an examination moderated the relationship between test anxiety and performance among 615 teenagers in the United Kingdom. Students were split into three groups of those taking a low-stakes, mock examination, those taking a mid-stakes, modular examination that is one of a series of tests taken over the final 18 months of schooling, and those taking a high-stakes, terminal examination. Putwain found that contrary to prediction, students taking the mock exam reported significantly greater test anxiety than students taking either the modular or terminal exam. This finding does not support the hypothesis that higher stakes are associated with a greater degree of perceived threat by children. However, Putwain argued that students taking the mock exam were experiencing their first exposure to this type of testing, whereas students taking the mid- and high- stakes exams had previous experience with testing (see Figure 1). Hence, the greater level of test anxiety found in the low-stakes group may have been an artifact of this study design.

Putwain (2008) also found that there were small but significant negative relationships between test anxiety, its components, and performance on all three exam types. The correlations between examination grade and the various measures of test anxiety including test anxiety, worry, and emotionality were *rs* = -0.19, -0.18, -0.14, respectively (Putwain). When students were dichotomized into low- and high-test anxious groups using test anxiety scores ± 1 standard deviation from the sample mean, high test-anxious students performed significantly lower than their low test-anxious peers on the mock and modular examinations. Alternatively, there was a small facilitating relationship between high test anxiety and examination performance on the terminal

examination. Putwain cautioned against the over-interpretation of these findings and called for replication of this study with a more robust methodology. In sum, this study raises important questions about the relationship between assessment stakes, test anxiety, and student performance that merit further study.

Indirect Studies of the Impact of Testing on Children

Jones and colleagues (1999) interviewed 236 teachers about their perceptions of the impact of high-stakes testing on students, themselves, and their instructional practices. Teachers were from North Carolina where a state accountability program that linked assessment results to employment decisions and financial incentives for teachers had been recently implemented. Teachers reported that instructional patterns had changed as a result of the accountability program, with more than 70% of teachers reporting spending more time practicing for end-of-grade tests and 80% of teachers reporting that they spent at least 20% of instructional time practicing for tests. Teachers also reported that these instructional changes and the testing program resulted in negative outcomes for children, with 61% of teachers reporting students felt more anxiety and 24% of teachers reporting that students were less confident. These reported impacts on children must be interpreted with considerable caution, however, since the authors did not directly assess how the assessment program affected students.

Another study of 708 teachers and 325 administrators in Florida found that teachers have greater concerns about the negative impact of high-stakes assessments on instructional practices and on students than administrators (Jones & Egley, 2006). While 46% of teachers reported that the testing program had negative effects on student and teacher motivation, only 27% of administrators reported this same belief. When these

same teachers responded to an open-ended question asking them to describe how the state assessment was impacting Florida in a positive or negative way, Jones and Egley (2004) described teachers' responses as being overwhelmingly negative, with teachers providing 89% negative reasons and only 8% positive reasons. Twenty-five percent of responses indicated that the testing program placed too much pressure on teachers and students, resulting in stress, anxiety, worry, or fear. However, the survey required teachers to decide if the state program took Florida in a positive or negative direction, leading teachers to focus on either positive *or* negative outcomes rather than providing a more complete description of both positive *and* negative outcomes of the program.

Similarly, Barksdale-Ladd and Thomas (2000) reported that teachers and parents feel that high-stakes assessments have negative impacts on students and teachers. In interviews with 59 teachers and 20 parents in two states with high-stakes assessment programs tied to monetary consequences for schools and grade retention for students, parents and teachers reported that the testing programs resulted in instructional changes and student and teacher stress and anxiety. Teachers described feeling pressure to ensure that students perform well, with some estimating that two to three months of the school year was focused on preparing for the test. Teachers described concern that their own stress and anxiety about the assessment program was transferred to students and that the children worried about testing and its consequences. Similarly, although parents reported more positive beliefs about the uses of testing, they reported that children experienced high levels of anxiety and nervousness prior to the tests and they expressed disappointment after learning how they scored. These four reports suggest that teachers have a number of concerns about the negative impact of testing on students and teachers,

and they emphasize the need to study the impact of testing on teachers and students directly.

Additionally, in a national study of teachers' opinions about state testing programs, Abrams, Pedulla, and Madaus (2003) compared testing programs with high versus low stakes, with high-stakes being defined as tests linked to significant consequences such as accreditation, financial rewards, or school autonomy. Teachers reported that the stakes associated with testing programs significantly impacted instructional practices and student stress. For example, 43% of teachers in high-stakes testing states, as opposed to only 17% in low-stakes states reported that time spent on instruction in tested areas increased a great deal, with decreased time spent in class enrichment or extracurricular activities. Forty-one percent of teachers in high-stakes states versus 18% in low-stakes states reported that there is so much pressure for students to score highly on the tests that they have little time to teach anything not on the test. In addition, 80% of teachers in high-stakes states perceived that students were under intense pressure to perform well, with 35% reporting that students were extremely anxious about the state test. In low-stakes states, 49% of teachers perceived that students were under intense pressure to perform well, with 20% reporting that students were extremely anxious about the state test. Therefore, the stakes associated with state testing programs appear to be directly related to the impact that testing programs have on students and schools.

Alternatively, other reports suggest that educational stakeholders such as school counselors and parents do not perceive high-stakes testing to have pervasively negative impacts on students. Brown, Galassi, and Akos (2004) surveyed 280 North Carolinian

school counselors and found that while 65% of school counselors reported concern that testing resulted in student stress, anxiety, fear, or pressure, only 4% reported that students experienced lowered self-confidence or self-esteem and only 2% reported that students experienced feelings of failure or incompetence due to testing. Also, only 24% of school counselors reported that the program had no positive student outcomes, with the others reporting positive outcomes in the form of increased student focus on schoolwork or in increased accountability. These results suggest that while anxiety or stress may be present during the testing situation, school counselors perceive this distress as transitory and as not having lasting impacts on students.

Similarly, Mulvenon, Stegman, and Ritter (2005) reported that parents were generally supportive of standardized assessment programs. Only 24% of parents reported that testing was significantly or extremely stressful for their children and only 13% reported that they felt significant or extreme pressure to help their children score well, suggesting that in general parents do not project high levels of stress to their children about the testing process. Therefore, these two studies suggest that parents and school counselors may have more positive views of testing and fewer concerns that testing has negative psychological consequences for students than teachers. In turn, the ways that different adults perceive testing and communicate with students about testing may affect how students themselves respond to testing.

In sum, these indirect and direct reports by students, teachers, parents, administrators, and school counselors provide the basis for further study of the impact of testing programs on students. A number of studies suggest that state testing programs result in heightened student anxiety, stress, and lowered motivation (Abrams et al., 2003;

Barksdale-Ladd & Thomas, 2000; Jones & Egley, 2004; 2006; Jones et al., 1999; Triplett & Barksdale, 2005). Alternatively, Mulvenon and colleagues (2001; 2005) found that most students felt positively about testing, did not experience stress as a result of testing, and their performance was not impaired by test anxiety. These studies provide important initial data about how various stakeholders perceive the impact of testing on students; however, there is a clear need to for additional studies that directly assess the impact of testing programs on children in light of the contradictory findings in the literature and the methodological weaknesses of previous studies.

Current Study

As illustrated by the literature review above, relatively little research directly examining test anxiety in children taking high-stakes tests has been conducted and little is known about the effect of test anxiety on student performance on high-stakes assessments. NCLB legislation has placed unprecedented weight on using students' test scores on statewide achievement assessments to evaluate and make decisions about public schools. Given meta-analytic data indicating test anxiety inversely affects student performance on achievement assessments, impairs academic performance in school, and results in lower educational attainment (Hembree, 1988), it is essential to examine the relationship between test anxiety and high-stakes testing. The existing literature predominantly used indirect methodology to study the impact of testing on students and there is a need for studies that directly examine the impact of testing on students. This dissertation study critically examined the impact that test anxiety has on students taking high-stakes achievement assessments and it filled gaps in the research literature on this topic. This goal was achieved by directly measuring students' test anxiety in relation to

high-stakes testing and classroom testing and examining the relationship between test anxiety, psychosocial variables, and test performance.

The subjects in this study were comprised of school children taking state-mandated achievement assessments required by NCLB. Young students, who have traditionally not been subject to standardized testing, have been most impacted by the annual testing requirements of NCLB. Therefore, students in elementary schools were the focus of this study in order to add to the limited literature on the expression and the impact of test anxiety among young children. Figure 2 presents the current study's methodological organization. The current study assessed the key biopsychosocial factors contained in this study's conceptual framework by including the measurement of overall test anxiety, worry, emotionality, and test behavior; academic achievement; test performance; and multiple psychosocial factors. Using this methodology, this dissertation study significantly expands upon the current test anxiety literature by critically examining how young children respond to high-stakes achievement testing.

While additional person-specific variables, including prior testing experience, trait anxiety, achievement need, social-emotional functioning, self-efficacy, social skills, and perception of the task, and testing/environmental variables, including testing atmosphere, examiner characteristics, task difficulty, time constraints, and mode of administration, are hypothesized to affect student test anxiety (see Figure 1), these factors were not included in the current study. The additional person-specific variables were not included due to methodological constraints (e.g., time and financial). The testing/environmental variables were not included in the study since the mandated achievement testing administration is standardized, using identical assessments, instructions, and administration dates.

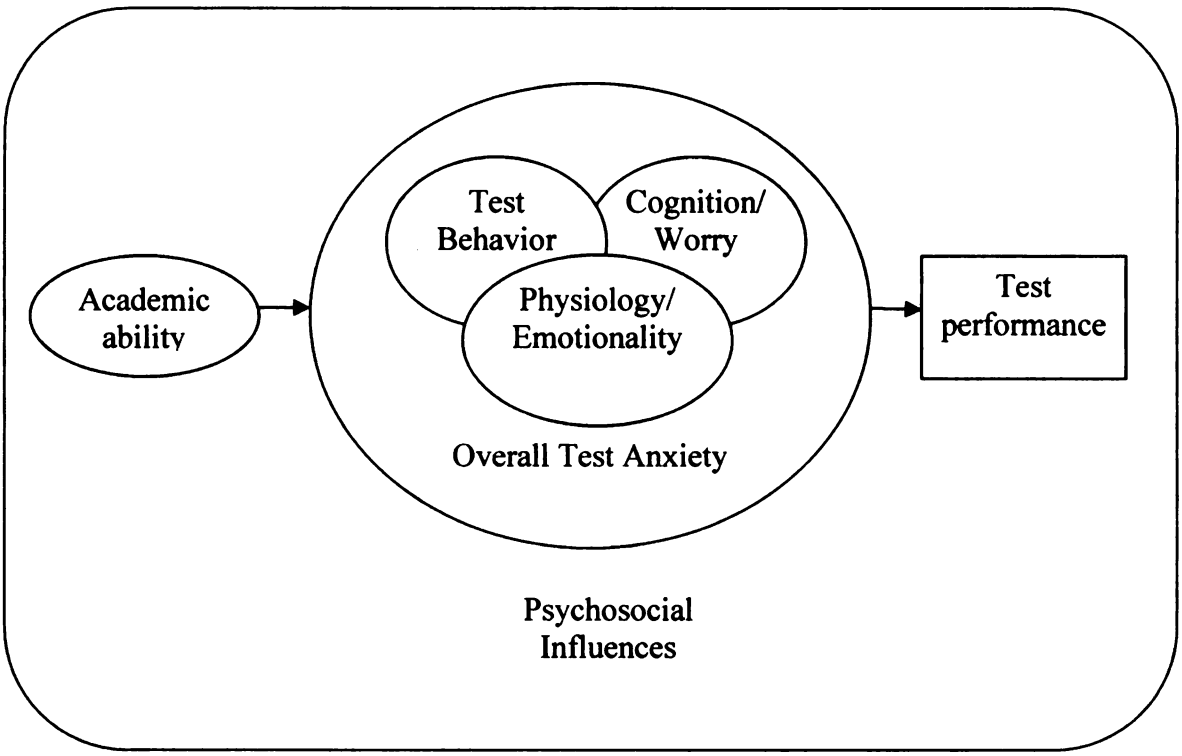


Figure 2: Methodological Design of the Current Study

Therefore, variability in testing/environmental variables is minimized in this study.

Research Questions and Hypotheses

Question One. Are there differences in test anxiety levels and rates associated with high-stakes testing and classroom testing? The first research question examined the differences in the level of reported test anxiety experienced by students taking state-mandated educational assessments as well as classroom tests. Additionally, differences in rates of low, moderate, and high test anxiety across the test conditions were examined. Previous test anxiety prevalence estimates have ranged from as low as 10 to as high as 41% (King & Ollendick, 1989; Turner et al., 1993). Although these rates have been determined using different operational criteria, the rates all include students who respond to testing situations with heightened anxiety and worry. Studies indicate that when students perceive assessments to have high personal relevance or when assessments are presented as measures of ability, they experience greater anxiety (Zeidner, 1998). The No Child Left Behind Act of 2001 has increased the frequency and the stakes of state educational assessments by linking essential funding, administrative policies, and decisions to student performance data (U.S. Congress, 2002). This federal mandate increased the direct relevance that state assessments have for educators, parents, communities, and schools. In turn, the high-stakes associated with state assessments may increase the prevalence of test anxiety in children. The researchers hypothesized that test anxiety levels among students would be significantly greater for high-stakes assessments than classroom testing. Additionally, the differences in rates of low, moderate, and high test anxiety across testing conditions were explored using different classification systems for test anxiety. The researchers hypothesized that rates of high test anxiety would be

significantly greater for high-stakes assessments than classroom testing.

Question Two. Are there differences in classroom teachers' perceptions of students' test anxiety responses and performance in relation to high-stakes testing and classroom testing? Are there differences in classroom teachers' anxious responses to high-stakes testing and classroom testing? Previous studies suggest that teachers believe high-stakes testing programs result in heightened student and teacher anxiety and stress (Abrams et al., 2003; Barksdale-Ladd & Thomas, 2000; Jones & Egley, 2004; 2006; Jones et al., 1999). However, differences in teacher perceptions across different test conditions have not been previously examined. The second research question examined the differences in teachers' perceptions of both their own and students' anxiety responses to testing and associated test performance across testing conditions. The researchers hypothesized that teachers would perceive students to have more test anxiety in relation to high-stakes assessments than classroom testing. Similarly, the researchers hypothesized that teachers would report more anxiety themselves about high-stakes assessments than classroom testing. Finally, the researchers hypothesized that teachers would report that student performance was more impaired on high-stakes assessments than classroom tests.

Question Three. Do students with different test anxiety classifications perform differently on high-stakes testing? Does test anxiety predict testing performance? Are there other predictors of test performance? Perhaps most central to the argument behind NCLB is the assumption that student test scores reliably and accurately represent students' educational experiences and knowledge. The use of assessments to make decisions about school funding, employment, and curriculum is based on the assumption

that the tests are indicative of the quality of the educational services being provided to students. However, research suggests that student performance on standardized assessments is related to student test anxiety (Hembree, 1988; Zeidner, 1998). The researchers hypothesized that test performance would differ on the basis of students' test anxiety level. Additionally, the researchers hypothesized that even after controlling for student achievement using academic indicators and student demographic characteristics, test anxiety would be a significant predictor of performance on the state assessment. Student achievement needs to be controlled for in order to account for pre-existing individual differences among students that also affect test scores. The test anxiety literature also suggests that impaired reading scores are more strongly associated with test anxiety than other academic subject areas (Sarason et al., 1960), leading the researchers to hypothesize that test anxiety would be more strongly related to impaired reading scores than mathematics scores. If test anxiety is found to be significantly related to test outcomes on the high-stakes assessment, this relationship may call into question the validity of making high-stakes decisions based on a sole indicator of student performance.

Question Four. Do students with different demographic characteristics report significant different levels test anxiety? How predictive are student demographic characteristics of test anxiety across the testing conditions? The test anxiety literature suggests that numerous demographic variables are associated with test anxiety. However, these variables have yet to be examined in relation to high-stakes tests. This research question examined whether or not, and to what degree, demographic characteristics predicted test anxiety for both classroom testing situations and high-stakes

testing situations. Since NCLB requires that student performance data be reported in groups that are disaggregated by social characteristics, including race, economic status, and disability status, among others, it is essential to examine if these variables are associated with test anxiety. As a result of student data being reported separately each of these student subgroups, greater pressure may be placed on these students to perform well on assessments, resulting in greater student anxiety and worry (Steele & Aronson, 1995). If certain groups of children respond to the assessment in a way that hinders their ability to perform, the use of their scores to make high-stakes decisions needs to be critically reexamined. Based on theory and previous findings, the researchers in the current study hypothesized that students receiving special education services, female students, older students, racial minority students, and economically disadvantaged students would experience greater test anxiety than other students on the high-stakes assessment (Hembree, 1988; Putwain, 2007). Additionally, the researchers in the current study hypothesized that female students, older students, racial minority students, and economically disadvantaged students would report greater levels of test anxiety on classroom tests (Hembree).

Question Five. What is the strength of the relationship between the Children's Test Anxiety Scale, the Behavior Assessment Scale for Children-2nd Edition-Test Anxiety Subscale, and the Test Anxiety Scale for Children within testing conditions? How closely do classroom testing test anxiety scores relate to high-stakes assessment test anxiety across measures? The most commonly used measure of test anxiety among children is the Test Anxiety Scale for Children (TASC, Sarason et al., 1960). It has been used extensively in studying test anxiety among young school children. However, questions

about the developmental appropriateness of the scale have been raised because the items are wordy and include complex sentence structures. Although designed for children as young as first grade, the scale's linguistic complexity raises the question of whether or not children are able to fully understand the questions. Therefore, it is important to examine more contemporary scales of test anxiety to determine their construct validity in relation to this established scale. The Children's Test Anxiety Scale (CTAS; Wren & Benson, 2004) and the Behavior Assessment Scale for Children, Second Edition, Test Anxiety subscale (BASC-2-TA; Reynolds & Kamphaus, 2004) are both recently developed measures of test anxiety. The relationship between the CTAS, BASC-2-TA, and TASC scales were examined in order to provide preliminary data about the newer scales' construct validity. On the basis of the scales' similar theoretical constructs, the researchers hypothesized that student's scores on the three scales would correlate moderately to highly with one another (rs $\geq$ 0.60). Additionally, the researchers hypothesized that student test anxiety in relation to classroom testing would be strongly correlated with student test anxiety in relation to high-stakes testing across all measures.

CHAPTER 3

METHOD

Subjects

School children in grades three through five in three elementary schools from a suburban school district in the state of Michigan were invited to participate in the study. This age population was selected based on the design of No Child Left Behind (NCLB) testing programs which require annual achievement testing beginning in grade three as well as data suggesting that test anxiety stability becomes significantly greater by late-elementary school (Hill & Sarason, 1966). All students taking the Michigan Educational Achievement Program (MEAP) were eligible for the study. Students who took the alternative assessment were not included; NCLB allows up to one percent of students to take an alternative assessment due to severe cognitive disability. Decisions about the alternative assessment are made by a student's Individualized Education Program Team when, due to severe cognitive impairment, the MEAP, even with accommodations, is deemed to be an inappropriate assessment.

A total of 617 children from 25 classrooms were invited to participate in the study. Parental consent forms permitting children to participate in the study were returned for 363 children. Among the three participating schools, 121 children received parental consent to participate in school A (54% of invited), 65 in school B (48% of invited), and 177 in school C (69% of invited). This overall 58% rate of student participation is excellent for survey research (Hopkins & Gullickson, 1992).

Following survey administrations, the total number of students who completed at least one survey was 335 students (92% of those with consent for an overall participation rate of 54% among all invited students). Children who did not participate in at least one

of the surveys were dropped from the study. One hundred and nineteen children completed surveys in school A (98% of those with consent). 50 in school B (78% of those with consent), and 166 in school C (94% of those with consent). Differences in participation rates across schools were related to student absences from school and student participation in alternative activities at the time of survey administration. Among the 25 teachers invited to complete teacher questionnaires, 88% participated in the first questionnaire and 60% participated in the second questionnaire. Differences in teacher completion of questionnaires may have been related to the increased length of the second questionnaire.

Table 1 summarizes the demographic characteristics of student participants. In some cases, the sample N does not total to 335, indicating that some demographic information was unavailable for participants. There were 151 males and 180 female students who ranged in age from 7 to 11 ($M = 9.10$ years, $SD = 0.92$), with 10.1% receiving special education services. On average, parents completed between one and four years of college and were from the middle class (SES: $M = 2.10$, $SD = 0.86$). The racial/ethnic makeup of the overall sample was 82.1% Caucasian, 4.5% African American, 7.2%, Hispanic or Latino, 3.0% Asian, and 2.4% Other.

The demographic characteristics of the total sample were examined to determine if participants in the three schools were representative of the actual schools' populations and if the samples differed significantly across schools. Differences in demographic characteristics between the overall population in the three schools and the overall sample were examined with Chi-square tests, using a Bonferonni correction of 0.125 for four comparisons. Chi-square goodness of fit tests indicated that there were no significant

Table 1

Demographic Characteristics of Student Participants

	Sample		Population			
Demographic Characteristic	N	%	N	%	*F*	*p*
Grade					1.41	.49
3	119	35.5	203	32.9		
4	124	37.0	229	37.1		
5	92	27.5	185	30.0		
Gender					5.85	.02
Male	151	45.1	682	52.3		
Female	180	53.7	623	47.7		
Race					6.71	.08
Caucasian	275	82.1	1098	84.1		
African American	15	4.5	81	6.2		
Latino/Hispanic	24	7.2	63	4.8		
Asian	10	3.0	55	4.2		
Other	8	2.4				
Educational Verification					3.15	.08
General Education	289	86.3	1123	86.1		
Special Education	34	10.1	182	13.9		
Total N	335		1305			

differences between the sample by grade, sex, ethnicity, or educational verification and the overall population of the three schools.

Additionally, differences in demographic characteristics between the overall school populations and the school samples were examined using Chi-square tests, using a Bonferonni correction of 0.17 for three comparisons. Chi-square goodness of fit tests indicated that there were no differences between the racial/ethnic or gender makeup of study samples from schools A, B, and C and the actual population of those schools. Similarly, there was no difference between the rate of students receiving special education services in the current study sample from schools A or B and the population of schools A and B. In school C, significantly fewer students receiving special education services were included in the current sample (10%) as compared to the school population proportion of 18%, χ^2 (1, n = 165) = 6.18, p = .01.

Table 2 summarizes the demographic characteristics of student participants for each of the participating schools. Differences in demographic characteristics across the samples from the three schools were also examined using analyses of variance and Chi-square tests, with a Bonferonni correction of $p \leq 0.008$ for six comparisons. A one-way between-groups analysis of variance was conducted to explore whether or not there were differences in parental education by school. There was a significant difference in parental education levels for the three schools: F (2, 330) = 6.70, p = .001. Parents in school B reported less education than parents in schools A and C. A one-way between-groups analysis of variance was also conducted to explore whether or not there were differences in family socioeconomic status by school. There was a significant difference in socioeconomic status levels for the three schools: F (2, 305) = 5.29, p = .006. Parents in

Table 2

Demographic Characteristics of Student Participants by School

Characteristic	School A		School B		School C			
	M	SD	M	SD	M	SD	*F*	*p*
Parent Education[a]	5.53	.77	5.16	.74	5.63	.83	6.70	.001
SES Code[a]	2.06	.82	2.46	.80	2.01	.88	5.29	.006
	N	%	N	%	N	%	χ^2	*p*
Grade							1.61	.81
3	43	36.4	20	39.2	56	33.7		
4	44	37.3	20	39.2	60	36.1		
5	31	26.3	11	21.6	50	30.1		
Gender							.35	.84
Male	51	44.0	24	49.0	76	45.8		
Female	65	56.0	25	51.0	90	54.2		
Race							5.37	.07
Caucasian	96	82.1	47	94.0	132	80.0		
Non-Caucasian	21	17.9	3	6.0	33	20.0		
Educational Verification							1.33	.52
General Education	104	91.2	40	85.1	145	89.5		
Special Education	10	8.8	7	14.9	17	10.5		

[a]Parental Education: 2: 9th grade or less, 3: 10th or 11th grade , 4: High School Graduate, 5: Partial College , 6: College Graduation, 7: Graduate Training; SES: Hollingshead Socioeconomic Codes: 1: High Class, 2: Middle Class, 3: Working Class, 4 and 5: Low Class

school B reported lower socioeconomic status than parents in schools A and C. Chi-square tests for independence indicated no significant associations between school and grade: χ^2 (4, n = 335) = 1.61, p = .81, school and gender: χ^2 (2, n = 331) = .35, p = .84, school and race: χ^2 (2, n = 332) = 5.37, p = .07, and school and educational status: χ^2 (2, n = 323) = 1.33, p = .52. Based on these analyses, which indicate that with the exception of socioeconomic status and parental education, participants' demographic characteristics did not differ significantly across schools, and the analyses indicating that the samples from each school were representative of the schools' actual populations, the data across all schools were combined for all subsequent data analyses.

Measures

Two survey packets were developed for use in this study. The first survey was made up of the Children's Test Anxiety Scale (CTAS) and the Behavior Assessment Scale for Children, Second Edition, Test Anxiety subscale (BASC-2-TA). The second survey contained the original measures and also included the Test Anxiety Scale for Children. The following sections describe the test anxiety measures contained in the survey packets:

Children's Test Anxiety Scale

The Children's Test Anxiety Scale (CTAS) is a 30-item self-report scale designed to measure test anxiety in children in grades three through six (Wren & Benson, 2004). The scale includes items that assess children's thoughts, off-task behaviors, and physiological reactions to testing. Each item is scored on a four-point Likert scale that ranges from *almost never* to *almost always*. Scores on the CTAS range from 30 to 120. The scale was constructed and validated on a sample of 261 children. Internal consistency

estimates for the CTAS were satisfactory for the overall scale (0.92), and for off-task behavior (0.76), physiological reactions (0.82), and thoughts (0.89) subscales. Administration time ranged from 5 to 12 minutes. Means and standard deviations for the total sample were 61.97 (16.49) on the overall scale, 16.89 (5.14) on the off-task behavior subscale, 15.96 (5.63) on the physiological reactions subscale, and 29.12 (8.79) on the thoughts subscale. Girls reported significantly greater overall test anxiety than boys, with mean scores of 65.78 (16.93) versus 57.71 (14.94). No classification system for clinical levels of test anxiety on the CTAS has been developed, although the test developer recommends classifying scores falling one standard deviation above the mean of the standardization sample as high test-anxious and scores falling one standard deviation below the mean as low test-anxious (Wren, personal communication, March 14, 2008).

Behavior Assessment Scale for Children, Second Edition, Test Anxiety Subscale

The test anxiety content subscale of the Behavior Assessment Scale for Children, Second Edition (BASC-2-TA) is made up of seven items on the self-report of personality (Reynolds & Kamphaus, 2004). The subscale measures a child's propensity for irrational worry and fear regarding test-taking. Scores on the subscale range from 0 to 18. The scale has been normed for children aged 12 to 18. In a sample of 884 12-14 year olds in the normative sample, the test anxiety subscale had an internal consistency alpha of 0.81. Means and standard deviations for the sample were 6.74 (3.67) when the sample was combined, 7.52 (3.61) for females, and 5.95 (3.57) for males. The BASC-2 classifies scores falling one standard deviation above the mean of the standardization sample as At-Risk and scores falling two standard deviations above the mean as Clinically Significant. No normative data for children aged 7 to 11 is available for the BASC-2-TA.

Test Anxiety Scale for Children

The Test Anxiety Scale for Children (TASC) is a 30-item self-report scale for children in grades one or higher (Sarason et al., 1960). The scale is designed for either researcher or teacher administration, with no differences in outcomes observed between these two conditions (Sarason et al.). The scale's construct validity has been demonstrated through studies demonstrating a significant correlation between the TASC and teacher ratings of student anxiety as well as consistently positive correlations (ranging from 0.47 to 0.69) found between the TASC and the General Anxiety Scale for Children (GASC) in American school children (Sarason et al.). The scale has a four-month test-retest reliability of 0.67 (Sarason et al.). Children who score above or below established cutpoints are classified as being high or low test-anxious. Girls with scores above 16 and boys with scores above 12 are considered to be high test-anxious. Girls with scores below 10 and boys with scores below 7 are considered to be low test-anxious (Beidel & Turner, 1988).

In addition to the test anxiety surveys, student demographic, academic achievement, and MEAP test performance data were collected. The following sections describe each area of data collection.

Demographic Information Form

Demographic information was collected through a form that was included with the parental consent form. Parents reported their child's date of birth, age, grade, sex, ethnicity, and educational verification (i.e., general education versus special education). Parents also reported on their educational attainment and employment status. Socioeconomic status (SES) was calculated using the Hollingshead Four Factor Index of

Social Status (Hollingshead, 1975). The Demographic Information Form is included in Appendix A.

Academic Performance Form

In order to examine whether or not student performance on the state assessment was affected by test anxiety, it was necessary to control for students' prior achievement. Some previous researchers have attempted to do this using student scores on previous standardized assessments (e.g., Mulvenon et al., 2001; 2005); however, these scores are likely to be similarly affected by test anxiety. Therefore, students' course grades were used as proxies for student achievement. Although course grades have some stakes associated with them, the grades children receive are based on a diverse set of course assignments, increasing the likelihood that the grades are accurate representations of students' true achievement. Student grades in English language arts, mathematics, and science were collected for the first grading period of the academic year. Science grades were also collected for current fifth grade students who took the MEAP subject test in science. The Academic Performance Form is included in Appendix B.

Michigan Educational Assessment Program Data

The Michigan Educational Assessment Program (MEAP) is the state accountability testing program used to comply with the requirements of the No Child Left Behind Act of 2001 (NCLB) in the state of Michigan. The MEAP is a criterion-referenced test that measures students' skills in the core content areas of mathematics, science, social studies, and English language arts. Testing occurred during the last three weeks of October 2008. Third, fourth, and fifth grade students took the MEAP assessment in the areas of Reading and Writing (English Language Arts) and

Mathematics. Fifth grade students also took the MEAP Science content test. Test data were reported through class grade roster reports and entered into the database by the researchers.

Scale scores are reported for each content area. For third grade students, a score of 300 is the lowest score in the proficient range. The same pattern is true for all grades, i.e., 400 for fourth grade, 500 for fifth grade, etc. are the lowest scores in the proficient range for these grades; hence, scale scores are not comparable across grades. A third grade student who scores 400 is not performing at a fourth grade level. Scores are also classified as ranging from Level 1 to Level 4. Level 1 indicates an Advanced Student, Level 2 indicates a Proficient student, Level 3 indicates a Partially Proficient student, and Level 4 indicates a Not Proficient student (State of Michigan, 2007). Additionally, students are classified as performing in the low, middle, or high range of each Level. Thus, student performance level is reported on a 12-point scale, with students classified as being low, middle, or high non-proficient (1 - 3), low, middle or high partially proficient (4 - 6), low, middle or high proficient (7 - 9), or low, middle, or high advanced (10 - 12).

Teacher Perception Questionnaires

Teacher perceptions of student test anxiety, teacher anxiety, and student performance were collected through two brief teacher questionnaires. Teachers reported on their perceptions of students' anxiety before and during testing conditions as well as their own anxiety related to student performance. Anxiety was rated on a 4-point Likert scale ranging from "not at all" to "very" anxious, with higher scores indicating higher anxiety. Finally, teachers reported on their perceptions of how students performed during

the two types of testing conditions. Performance was rated on a 5-point Likert scale ranging from "very much below their ability" to "very much above their ability," with higher scores indicating higher performance. The Teacher Questionnaires are included in Appendix C and D.

Operational Definitions of Test Anxiety

Students were classified as having low, moderate, or high test anxiety using three definitions of test anxiety. Three definitions of test anxiety were used in order to examine whether or not there were classification differences between the TASC, a frequently used and well-established measure of test anxiety, and the CTAS, a contemporary measure of test anxiety based on a biopsychosocial model. No test anxiety classifications were made using students BASC-2-TA scores due to age differences between the current sample and the normative sample for the BASC-2-TA.

Definition One

The first test anxiety definition was based on students' scores on the Children's Test Anxiety Scale, which examines behavioral, cognitive, and physiological aspects of test anxiety. Students were classified as having high test anxiety if their overall CTAS score was one standard deviation above the mean of the standardization sample (78.46) and as having low test anxiety if their CTAS score was one standard deviation below the mean (45.48, Wren & Benson, 2004; Wren, personal communication, March 14, 2008). Students whose test anxiety scores fall within one standard deviation of the mean were classified as having moderate test anxiety.

Definition Two

The second test anxiety definition was based on students' scores on the Test

Anxiety Scale for Children, which was developed by Sarason and colleagues in 1960, and has been the most extensively used tool to measure childhood test anxiety. Cutpoints established by Beidel and Turner (1988) were used to determine students' test anxiety classification. Girls with scores above 16 and boys with scores above 12 were classified as high test-anxious. Girls with scores below 10 and boys with scores below 7 were classified as low test-anxious. Girls with scores between 10 and 16 and boys with scores between 7 and 12 were classified as having moderate test anxiety.

Definition Three

The third and most conservative test anxiety definition was based on students' scores on both the Children's Test Anxiety Scale and the Test Anxiety Scale for Children. Students classified as having high test anxiety by both Definitions One and Two were classified as high test-anxious and students classified as having low test anxiety by both Definitions One and Two were classified as being low test-anxious. Students classified as having moderate test anxiety by both Definitions One and Two were classified as moderately test-anxious. Additionally, students who did not have stable test anxiety classifications across the Definitions One and Two were classified as having moderate test anxiety.

Data Collection Procedures

Recruitment, Consent, and Assent

The elementary schools where the study took place were selected by contacting principals and explaining the study and its purpose. Schools were selected as a convenience sample on the basis of the schools' interest and commitment to the study. Three schools in a suburban school district, with a total population of 617 students in

grades three through five, agreed to participate. The school district reported that it did not engage in any specific test preparation for the MEAP, although teachers are encouraged to format some tests using a multiple choice format so that students become familiar with this testing format. All schools in the study had a history of meeting Annual Yearly Progress (AYP) goals for the past five years except for one school, which did not meet AYP in 2007-2008 because it only assessed 94% of the students classified as economically disadvantaged on one portion of the MEAP and AYP standards require that 95% of students are assessed.

An informational letter was mailed home to the parents of children in the participating schools that described the study and its purpose, and extended an invitation for voluntary participation (see Appendix E). Consent forms were included in the mailing, along with privacy envelopes to return consent forms to the school (see Appendix F). Letters were sent to parents approximately three weeks prior to the data collection. Prior to data collection, children were also informed that they had the right to choose not to participate in the study. Students were asked to sign a statement of assent on the cover page of each survey packet that they completed (see Appendix G and H). Participating students received nominal items, including pencils, stickers, and erasers, following each survey.

MEAP Test Anxiety Data Collection

Prior to data collection, researchers read standardized instructions explaining how the students should complete the measures (see Appendix I). Researchers administered the first survey (see Appendix G) within 1 to 3 days after the completion of the Michigan Educational Assessment Program (MEAP) administration. Measures

included the adapted Children's Test Anxiety Scale and adapted Behavior Assessment Scale for Children Test Anxiety Subscale. The CTAS and the BASC-2-TA measures were adapted for use in the first survey by altering the word *test* to *MEAP* throughout survey items. Measures were randomly ordered by classroom in order to control for order effects.

Classroom Test Anxiety Data Collection

Researchers administered the second survey (see Appendix H) 28 days after the first survey. This survey focused on children's responses to classroom tests. The delay between surveys was designed to help students clearly differentiate between the two surveys' foci, the first being MEAP testing and the second being classroom testing. Measures on the second survey included the Children's Test Anxiety Scale, the Behavior Assessment Scale for Children- Second Edition Test Anxiety Subscale, and the Test Anxiety Scale for Children. Researchers read a set of standardized instructions explaining how the students should complete the measures and reminded students of all classroom tests they had taken in the past two weeks. Measures were randomly ordered by classroom in order to control for order effects.

Data Analysis

Data were entered and stored in a computer database that was password protected. No identifying information was recorded in the database. All raw data were examined for any errors in coding through examination of descriptive statistics and visual presentations of data. Twenty-four percent of all surveys (n= 151) were recoded to assess interrater reliability of subscale and total scale scores across all three test anxiety measures. High interrater reliability was found between coders, with 98% agreement. Following data-

checking procedures, statistical analyses were used to examine response patterns and relationships in the data.

Question One

Question One had two primary aims: (1) examining differences in reported test anxiety between testing conditions and (2) examining differences in how the three definitions of test anxiety classified children as low, moderate, or high test anxious.

Aim 1. Descriptive statistics of reported test anxiety in relation to the MEAP and classroom tests on the CTAS, BASC-2-TA, and TASC were examined. Wilcoxon Signed Rank tests were used to determine whether or not students reported significantly different levels of test anxiety for classroom testing and MEAP testing conditions. Nonparametric statistics were conducted because the distributions of test anxiety scores were not normally distributed. The distributions of scores on all test anxiety measures were positively skewed. Test anxiety was considered a continuous variable for the purpose of these analyses. A Bonferroni correction of $p \leq .01$ was used to correct for multiple comparisons. Effect sizes of scores that differed significantly across test types were calculated, $r = z / \sqrt{2n}$.

Aim 2. Chi-square goodness of fit analyses were conducted to examine whether or not students' test anxiety classifications (Low, Moderate, High) as determined by the three definitions of test anxiety differed significantly by testing condition (MEAP versus classroom tests). A Bonferroni correction of $p \leq .01$ was used to correct for multiple comparisons. Additional Chi-square goodness of fit analyses were conducted to examine whether or not students' test anxiety classifications (Low, Moderate, High) differed significantly across different definitions within the same testing condition.

Question Two

The primary aim of Question Two was to examine differences in teacher perceptions of student test anxiety, teacher anxiety, and student performance between testing conditions. Independent-samples t-tests were conducted to examine differences because teacher perception data were collected anonymously and it was not possible to pair teacher questionnaires across testing conditions. When Levene's test for equality of variances indicated a violation of the assumption of equal variance, t-tests without the assumption of equal variance were examined.

Question Three

Question Three had two primary aims: (1) examining differences in MEAP test performance across students identified as having low, moderate, and high test anxiety and (2) examining predictors of high-stakes test performance.

Aim 1. Descriptive statistics of test performance on the English Language Arts and Math portions of the MEAP were examined and one-way between-groups analyses of variance with Tukey HSD post-hoc tests were used to determine whether or not students with different test anxiety classifications performed differently on the MEAP. For the purpose of this analysis, student performance level on the MEAP was used so that student performance data across all three grades could be combined. Effect sizes of performance levels that differed significantly across test anxiety classifications were calculated, eta squared = (Sum of squares between-groups) / (Total Sum of Squares).

Aim 2. Hierarchical multiple regression analyses were used to assess the ability of students' test anxiety scores to predict test performance on the MEAP. Student scholastic achievement was controlled for by including students' academic grades in each

regression analysis. Demographic variables, including student sex, SES, educational verification, and ethnicity, were also included in the regression analyses to examine the influence of these variables on student performance. Test anxiety was considered a continuous variable for the purpose of these analyses. Separate analyses were conducted for each grade level because student academic grades and MEAP scores were reported using different scales across grades three, four, and five.

Question Four

Question Four had two primary aims: (1) examining if there were systematic differences between students' test anxiety scores when students were grouped by different demographic variables and (2) examining if demographic variables significantly predicted test anxiety measure scores.

Aim 1. Nonparametric statistics were used to examine if there were significant differences in students' test anxiety scores as a result of differences in students by demographic variables. Nonparametric analyses were conducted because the distributions of test anxiety scores across all measures were not normally distributed. Test anxiety was considered a continuous variable for the purpose of these analyses. Mann-Whitney U tests were used to examine differences in dichotomous demographic variables, including sex, ethnicity, and educational verification. Kruskal-Wallis Tests were used to examine differences in demographic variables with three or more categories, including socioeconomic status and grade. A Bonferroni correction of $p \leq .01$ was used to correct for multiple comparisons. Effect sizes of scores that differed significantly across test types were calculated, $r = z / \sqrt{n}$.

Aim 2. Hierarchical multiple regression analyses were used to assess the ability of

demographic variables, including student grade, sex, SES, educational verification, and ethnicity, to predict test anxiety. Data transformations of test anxiety scores were completed to account for the non-normal distribution of survey data, which was positively skewed. Test anxiety was considered a continuous variable for the purpose of these analyses. The relationship between demographic variables and test anxiety were examined in relation to both MEAP and classroom testing conditions.

Question 5

Question Five had two primary aims: (1) examining the relationship between the different test anxiety measures within testing conditions and (2) examining the relationship between the measures across the two testing conditions.

Aim 1. Correlation analyses were conducted to examine the relationship between test anxiety scores on the CTAS, BASC-2-TA, and TASC for both testing conditions. Nonparametric, two-tailed Spearman rank order correlation (rho) analyses were used because children's reported test anxiety across measures was not normally distributed. Test anxiety was considered a continuous variable for the purpose of these analyses. Correlations were considered to be strong if $r_s \geq .60$, moderate if $.40 \leq r_s \leq .59$, and low if $r_s \leq .40$. Additionally, a Kappa Measure of Agreement was calculated for children classified using the TASC, the "gold standard" measure of test anxiety and the CTAS, a more contemporary measure of test anxiety designed to assess children's psychological, behavioral, and physiological response to testing, to examine how similarly the two measures' classified students as anxious or non-anxious.

Aim 2. Correlation analyses were conducted to examine the relationship between children's scores on test anxiety measures across testing conditions. Nonparametric, two-

tailed Spearman rank order correlation (rho) analyses were used because children's reported test anxiety across measures was not normally distributed. Test anxiety was considered a continuous variable for the purpose of these analyses. Analyses were conducted for the three different test anxiety scales and the subscales of the CTAS. Correlations between test conditions were considered to be strong if $r_s \geq .60$, moderate if $.40 \leq r_s \leq .59$, and low if $r_s \leq .40$.

CHAPTER 4

RESULTS

Question One

Aim 1

Wilcoxon Signed Rank Tests were used to examine the differences between test anxiety scores reported for MEAP and classroom tests across measures using a Bonferroni correction of $p \leq .01$. Table 3 summarizes the differences identified between reported test anxiety on the MEAP assessment and classroom tests across measures. There was a statistically significant difference between students' reported test anxiety on the CTAS-Total scale across the MEAP and classroom tests, $z = -5.04$, $p < .0005$, with a small effect size ($r = -0.21$). Students reported significantly greater test anxiety on the CTAS-Total scale in relation to the MEAP assessment than in relation to classroom tests ($Md = 52$ versus $Md = 47$, respectively). There was also a statistically significant difference in students' BASC-2-TA test anxiety scores between the MEAP and classroom tests, $z = -2.47$, $p = .01$, with a small effect size ($r = -0.10$). Students reported significantly greater test anxiety on the BASC-2-TA in relation to the MEAP than in relation to classroom tests ($Md = 4$ versus $Md = 4$, respectively).

Wilcoxon Signed Rank tests also identified significant differences between reported test anxiety on two CTAS subscales in relation to the MEAP and classroom tests. There was a statistically significant difference in students' CTAS-Thought subscale scores between the MEAP and classroom tests, $z = -4.87$, $p < .0005$, with a small effect size ($r = -0.20$). Students reported significantly greater test anxiety on the CTAS-Thought subscale in relation to the MEAP than in relation to classroom tests ($Md = 24$ versus Md

Table 3

Test Anxiety Scores on the MEAP Assessment and Classroom Testing

	MEAP Test[a]				Classroom Tests						
Measure[a]	*n*	*Md*	*M*	*SD*	*n*	*Md*	*M*	*SD*	*n*	*z*	*p*[b]
CTAS											
Total	305	52	55.97	16.33	318	47	52.21	18.21	292	-5.04	< .0005
Thoughts	307	24	25.36	8.12	320	21	23.24	9.24	294	-4.87	< .0005
Off-task	306	14	15.31	5.25	324	14	15.24	5.91	295	-0.08	0.93
Autonomic	307	14	15.35	5.96	322	12	13.80	5.75	295	-5.87	< .0005
BASC-2-TA	309	4	4.99	4.06	322	4	4.42	3.90	297	-2.47	0.01
TASC					321		8.49	7.01			

[a]MEAP: Michigan Educational Assessment Program; CTAS: Children's Test Anxiety Scale; BASC-2-TA: Behavioral Assessment Scale for Children, Second Edition, Test Anxiety Subscale; TASC: Test Anxiety Scale for Children

[b]Bonferonni correction of $p \leq .01$ for multiple comparisons.

= 21, respectively). There was also a statistically significant difference in students' CTAS-Autonomic subscale scores between the MEAP and classroom tests, $z = -5.87$, $p < .0005$, with a small effect size ($r = -0.24$). Students reported significantly greater test anxiety on the CTAS-Autonomic subscale in relation to the MEAP than in relation to classroom tests ($Md = 14$ versus $Md = 12$, respectively). There was no significant difference in students' CTAS-Off-task subscale scores between the MEAP and classroom tests, $z = -0.08$, $p = .93$.

Aim 2

Chi square goodness of fit tests were conducted to examine if there were significant differences between the proportions of students identified as low, moderate, or high test-anxious across the MEAP and classroom testing conditions using a Bonferroni correction of $p \leq .01$. Table 4 summarizes the differences identified between students' test anxiety classifications across the MEAP and classroom testing conditions. Using students' CTAS-Total scale scores to classify students, there was a significant difference in the proportions of students with low, moderate, and high levels of test anxiety in the classroom test condition as compared to the values of 32.2%, 58.6%, and 9.1%, respectively that were obtained in the MEAP condition, χ^2 (2, $n = 318$) = 27.90, $p < .0005$. More students reported low (143 vs. 102.5 expected) and high (34 vs. 29 expected) test anxiety and fewer students reported moderate test anxiety (141 vs. 186.4 expected) in the classroom testing condition than the MEAP condition.

Chi square goodness of fit tests also identified significant differences in students' test anxiety classifications using two of the CTAS subscales to classify students' test anxiety across the MEAP and classroom testing conditions. Using students' CTAS-

Table 4

Differences in the Number of Students Reporting Low, Moderate, and High Test Anxiety between MEAP and Classroom Testing Conditions by Definition

Measure[a]	MEAP Assessment Anxiety Level[a]			Classroom Testing Anxiety Level			χ^2	p[b]
	Low	Mod	High	Low	Mod	High		
Definition 1 (CTAS)								
Total	99	180	28	143	141	34	27.90	< .0005
Thoughts	99	180	28	153	136	31	38.36	< .0005
Off-task	85	190	31	102	178	44	8.08	.02
Autonomic	65	194	48	116	172	34	43.72	< .0005
Definition 2 (TASC)				169	82	66		
Definition 3 (stable CTAS/TASC)	82	195	13	119	145	26	42.24	< .0005

[a]MEAP: Michigan Educational Assessment Program; CTAS: Children's Test Anxiety Scale; TASC: Test Anxiety Scale for Children

[b]Bonferonni correction of $p \leq .01$ for multiple comparisons

Thought subscale scores to classify students, there was a significant difference in the proportions of students with low, moderate, and high levels of test anxiety in the classroom testing condition as compared to the values of 32.2%, 58.6%, and 9.1%, respectively that were obtained in the MEAP condition, χ^2 $(2, n = 320) = 38.36, p <$ .0005. More students reported low (153 vs. 103.2 expected) and high (31 vs. 29.2 expected) test anxiety and fewer students reported moderate test anxiety (136 vs. 187.6 expected) in the classroom testing condition than the MEAP condition. Using the CTAS-Autonomic subscale scores to classify students, there was also a significant difference in the proportions of students with low, moderate, and high levels of test anxiety in the classroom testing condition as compared to the values of 21.2%, 63.2%, and 15.6%, respectively that were obtained in the MEAP condition, χ^2 $(2, n = 322) = 43.72, p <$ .0005. More students reported low (116 vs. 68.2 expected) test anxiety and fewer students reported moderate (172 vs. 203.5 expected) and high (34 vs. 50.3 expected) test anxiety in the classroom testing condition than the MEAP condition. Using the CTAS-Off-task subscale scores to classify students, there was no significant difference in the proportions of students with low, moderate, and high levels of test anxiety in the classroom testing condition as compared to the values of 27.8%, 62.1%, and 10.1%, respectively that were obtained in the MEAP condition, χ^2 $(2, n = 324) = 8.08, p = .02$.

Finally, using students' stable CTAS-Total and TASC scores to determine student test anxiety levels, a Chi square goodness of fit test indicated that there was a significant difference in the proportions of students with low, moderate, and high levels of test anxiety in the classroom testing condition as compared to the value of 28.3%, 67.2%, and 4.5%, respectively that was obtained in the MEAP condition, χ^2 $(2, n = 290) = 42.24, p <$

.0005. More students reported low (119 vs. 82.1 expected) and high (26 vs. 13.0 expected) test anxiety and fewer students reported moderate (145 vs. 194.9 expected) test anxiety in the classroom testing condition than in the MEAP condition.

Figure 3 illustrates the differences identified in test anxiety classifications by Definition and test condition type, with the CTAS-Total classification proportions represented in the Definition 1 columns, the TASC classification proportions represented in the Definition 2 column, and the stable CTAS-Total and TASC classification proportions represented in the Definition 3 columns. Letters A-E below refer to the different columns represented on Figure 3. As discussed previously, students differed significantly in Definition 1 classifications across the MEAP (A) and classroom testing (B) conditions, $p < .0005$. Students also differed significantly in Definition 3 classifications across the MEAP (D) and classroom testing (E) conditions, $p < .0005$.

Additional Chi-square goodness of fit analyses were conducted to examine if there were significant differences in classification patterns across the different test anxiety classification definitions. First, differences were examined among students' test anxiety classifications on the MEAP assessment. There was a significant difference in the proportions of students with low, moderate, and high levels of test anxiety on the MEAP across the classifications established by Definition 1 and Definition 3. The students classified by the Definition 1 (A) significantly differed from the proportions of 28.3%, 67.2%, and 4.5%, respectively obtained by Definition 3 (D), χ^2 (2, n = 307) = 19.61, $p <$ .0005. More students reported low (99 vs. 86.9 expected) and high (28 vs. 13.8 expected) test anxiety and fewer students reported moderate test anxiety (180 vs. 206.3 expected) when classified using CTAS-Total scores versus stable CTAS-Total and TASC scores.

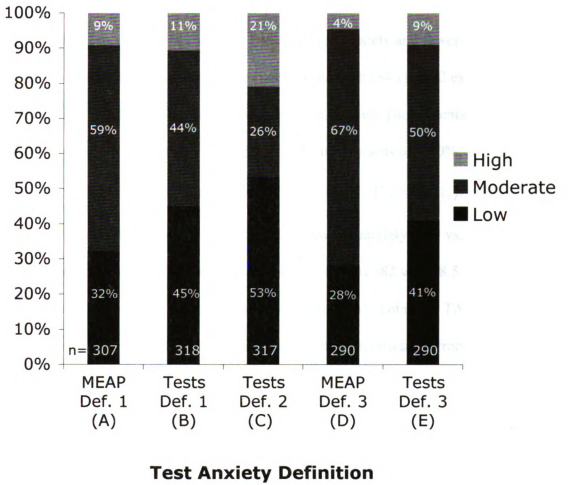


Figure 3: Rates of Test Anxiety by Classification Definitions

Next, differences between students' test anxiety classifications on the Classroom tests were examined across the different Definitions. There was a significant difference in the proportions of students identified with low, moderate, and high levels of test anxiety on classroom tests across the Definition 1 and Definition 2. The students classified by Definition 1 (B) significantly differed from the proportions of 53.3%, 25.9%, and 20.8%, respectively obtained by Definition 2 (C), χ^2 (2, n = 318) = 61.77, p < .0005. More children reported moderate (141 vs. 82.3 expected) test anxiety and fewer children reported low test anxiety (141 vs. 169.5 expected) and high (34 vs. 66.2 expected) when classified using CTAS-Total scores rather than TASC scores. The students classified by Definition 2 (C) also significantly differed from the proportions of 41.0%, 50.0%, and 9.0%, respectively obtained by Definition 3 (E), χ^2 (2, n = 317) = 98.21, p < .0005. More students reported high (66 vs. 28.4 expected) and low test anxiety (169 vs. 130.1 expected) test anxiety and fewer students reported moderate (82 vs. 158.5 expected) when classified using TASC scores rather than stable CTAS-Total and TASC scores. The students classified by Definition 1 (B) did not differ significantly from the proportions of 41.0%, 50.0%, and 9.0%, respectively obtained by Definition 3 (E), χ^2 (2, n = 318) = 4.29, p = .12.

Question Two

Teacher questionnaire data were analyzed to examine differences in teachers' perceptions of student anxiety, teacher anxiety, and student performance based on the testing condition. Table 5 summarizes differences in perceived anxiety and performance. There was a significant difference in teachers' perceptions of students' *anticipatory* test anxiety for the MEAP (M = 2.36, SD = .73) and classroom testing (M = 1.93, SD = .46); t

Table 5

Teacher Perceptions of Student and Teacher Test Anxiety by Testing Condition

Question	MEAP Assessment		Classroom Tests			
	M	*SD*	*M*	*SD*	*t*	*p*
1. How *anxious* were your students about [condition] before they took them?	2.36	.73	1.93	.46	2.21	.03
2. How *anxious* were your students about [condition] during the tests?	2.32	.65	2.00	.53	1.57	.13
3. How *anxious* were you about how well your students would do on [condition]?	2.59	.80	1.80	.68	3.15	.003
4. How well do you think your students performed on [condition]?	2.50	.58	2.80	.37	-1.93	.06

(34.85) = 2.21, p = .03 (two-tailed). There was no significant difference in teachers' perceptions of students' test anxiety *during* the MEAP (M = 2.32, SD = .65) or classroom testing (M = 2.00, SD = .53); t (35) = 1.57, p = .13 (two-tailed), although there was a trend toward rating students as more anxious during the MEAP. There was a significant difference in teachers' perceptions of their own anxiety about student performance on the MEAP (M = 2.59, SD = .80) and classroom tests (M = 1.80, SD = .68); t (35) = 3.15, p = .003 (two-tailed). There was no significant difference in teachers' perceptions of students' performance on the MEAP (M = 2.50, SD = .58) and classroom tests (M = 2.80, SD = .37); t (34.90) = -1.93, p = .06 (two-tailed), although there was a trend toward rating students as performing better on classroom tests than on the MEAP.

Question Three

Aim 1

One-way between-groups analyses of variance were conducted to examine whether or not students with different test anxiety classifications, as defined by CTAS-Total scores (Definition 1) and the TASC scores (Definition 2), performed significantly differently on the MEAP in the areas of English Language Arts, Math, and Science for students in grades three through five. First, analyses were conducted with students divided into low, moderate, and high test anxiety groups on the basis of their CTAS-Total scores. There was a significant difference between student performance level on the English Language Arts section of the MEAP for the three test anxiety groups: F (2, 297) = 7.07, p = .001. The effect size of the difference between groups was small, eta squared = .05. Post-hoc comparisons using the Tukey HSD indicated that the mean performance of the low test anxiety group (M = 8.81, SD = 1.19) was significantly greater than the

moderate test anxiety group (M= 8.17, SD = 1.47). The high test anxiety group (M= 8.20, SD = 1.35) did not differ significantly from either the moderate or low test anxiety groups. There was also a significant difference between student performance level on the Math section of the MEAP for the three test anxiety groups: F (2, 292) = 18.60, $p <$ .0005. The effect size of the difference between groups was moderate, eta squared = .11. Post-hoc comparisons indicated that the mean performance of the low test anxiety group (M = 9.86, SD = 1.26) was significantly greater than the moderate test anxiety group (M= 8.87, SD = 1.56) and the high test anxiety group (M= 8.24, SD = 1.76). The high test anxiety group did not differ significantly from the moderate test anxiety group.

Next, analyses were conducted with students divided into low, moderate, and high test anxiety groups on the basis of their TASC scores. There was a significant difference found between student performance level on the English Language Arts section of the MEAP for the three test anxiety groups: F (2, 308) = 6.24, p = .002. The effect size of the difference between groups was small, eta squared = .04. Post-hoc comparisons indicated that the mean performance of the low test anxiety group (M = 8.65, SD = 1.36) was significantly greater than moderate test anxiety group (M = 8.13, SD = 1.42) and the high test anxiety group (M= 8.05, SD = 1.47). The high test anxiety group did not differ significantly from the moderate test anxiety group. There was also a significant difference between student performance level on the Math section of the MEAP for the three test anxiety groups: F (2, 304) = 10.01, $p <$ .0005. The effect size of the difference between groups was moderate, eta squared = .06. Post-hoc comparisons indicated that the mean performance of the low test anxiety group (M = 9.53, SD = 1.41) was significantly greater than the moderate test anxiety group (M= 8.66, SD = 1.96) and the high test

anxiety group (M= 8.80, SD = 1.52). The high test anxiety group did not differ significantly from moderate test anxiety group. Table 6 summarizes the differences in MEAP performance level across low, moderate, and high test-anxious students.

Aim 2

Hierarchical multiple regression analyses were used to assess the ability of test anxiety in relation to the MEAP to predict MEAP performance. Analyses were conducted separately for students in each grade level. Students' academic achievement was controlled for by including a measure of scholastic achievement in each analysis. Additionally, the ability of student demographic characteristics to predict MEAP performance was also examined by including these variables in the regression analyses. Preliminary analyses were conducted for each regression analysis to ensure no violation of the assumptions of multiple regression. Tolerance and variance inflation factor scores (VIFs) were examined and no indications of multicollinearity were identified (Tols. > .10 and VIFs < 10). Normal probability plots of the regression standardized residuals and scatterplots of the standardized residuals were also examined and no violations of normality, linearity, homoscedasticity, or independence of residuals were identified for any of the regression analyses. The CTAS-Total scale was used as the only measure of test anxiety for these analyses and the BASC-2-TA scale was excluded to prevent multicollinearity between the two highly correlated variables (r_s = .71).

The first regression analysis for each grade examined the value of the CTAS-Total scale and student demographic characteristics in predicting test performance on the English Language Arts section of the MEAP when academic achievement was controlled for using students' reading grades. Academic achievement was entered in Step 1 and

Table 6

Differences in MEAP Performance Level by Test Anxiety Level[a]

MEAP Section	Low			Mod			High			*F*	*p*
	n	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*		
	CTAS-Total Classification										
English Language Arts	99	8.81	1.19	172	8.17	1.47	25	8.20	1.35	7.07	.001
Math	97	9.86	1.26	173	8.87	1.56	25	8.24	1.76	18.60	<.0005
	TASC Classification										
English Language Arts	167	8.65	1.36	78	8.13	1.42	64	8.05	1.49	6.24	.002
Math	166	9.53	1.41	77	8.66	1.96	64	8.80	1.52	10.01	<.0005

[a]MEAP: Michigan Educational Assessment Program; CTAS-Total: Children's Test Anxiety Scale Total score; TASC: Test Anxiety Scale for Children

included students' Reading Grade, the test anxiety measure was entered in Step 2 and included the CTAS-Total scale score, student sex was entered in Step 3 (female = 1, male = 0), socioeconomic status was dummy coded in Step 4: SES II (SES-II = 1, other = 0), SES-III (SES-III = 1, other = 0), and SES-IV-V (SES-IV-IV = 1, other = 0), educational verification was entered in Step 5: (special education = 1, general education = 0), and ethnicity was entered in Step 6: (Non-Caucasian = 1, Caucasian = 0). The second regression analysis for each grade examined the value of the CTAS-Total scale and student demographic characteristics in predicting test performance on the Math section of the MEAP when academic achievement was controlled for using students' math grade. The regression analysis was entered in the same manner as the previous analysis with the exception of entering students' math grades in Step 1.

Grade 3. The regression analysis of third grade students' English Language Arts MEAP scores as the dependent variable indicated that reading grade explained 36.0% of the variance in third graders' English Language Arts MEAP scores, $F(1, 94) = 52.98, p < .0005$. After the CTAS-Total was entered in the model, the total variance explained was 37.3%, R squared change = .013, F change $(1, 93) = 1.92, p = .17$, indicating that CTAS-Total did not contribute significantly to the model. After the sex was entered in the model, the total variance explained was 41.4%, R squared change = .040, F change $(1, 92) = 6.30, p = .01$. After SES was entered in the model, the total variance explained was 46.1%, R squared change = .048, F change $(3, 89) = 2.62, p = .06$. Educational verification and ethnicity variables did not add significantly to the model, $ps > .10$. The overall model was a significant predictor of English Language Arts scores on the MEAP, $F(8, 87) = 9.87, p < .0005$. In the final model, reading grade, CTAS-Total score, sex, and

SES-III variables were significant predictors of English Language Arts MEAP scores, with reading grade recording the highest beta value ($\beta = .50, p < .0005$) in comparison to the other variables, SES-III ($\beta = -.24, p = .02$), sex ($\beta = .21, p = .01$), and CTAS-Total ($\beta = -.17, p = .04$).

The regression analysis of third grade students' Math MEAP scores as the dependent variable indicated that math grade explained 9.7% of the variance in third graders' Math MEAP scores, $F(1, 94) = 10.09, p = .002$. After the CTAS-Total was entered in the model, the total variance explained was 20.4%, R squared change = .107, F change $(1, 93) = 12.47, p = .001$. After the educational verification was entered in the model, the total variance explained was 32.7%, R squared change = .094, F change (1, 88) = 12.29, $p = .001$. Sex, SES, and ethnicity variables did not add significantly to the model, $ps > .10$. The overall model was a significant predictor of MEAP Math scores, $F(8, 87) = 5.40, p < .0005$. In the final model, CTAS-Total and educational verification variables were the only significant predictors of Math MEAP scores, with CTAS-Total ($\beta = -.35, p < .0005$) recording a higher beta value than educational verification ($\beta = -.32, p = .002$). Tables 7 summarizes how predictive students' CTAS-Total scale scores, academic grades, and demographic characteristics were in accounting for variance in English Language Arts and Math MEAP scores for students in grade three.

Grade 4. The regression analysis of fourth grade students' English Language Arts MEAP scores as the dependent variable indicated that reading grade explained 14.1% of the variance in fourth graders' English Language Arts MEAP scores, $F(1, 104) = 17.09$, $p < .0005$. After the CTAS-Total was entered in the model, the total variance explained was 17.8%, R squared change = .037, F change $(1, 103) = 4.67, p = .033$. After SES was

Table 7

Multiple Regression Of Predictors of MEAP Scores in Grade 3

	MEAP					
	English Language Arts			Math		
Additive Model	β^a	ΔR^2	F-Δ	β^a	ΔR^2	F-Δ
+Achievement		.360	52.98***		.097	10.09**
Reading Grade	.50***					
Math Grade				.17†		
+CTAS-Total	-.17*	.013	1.92	-.35***	.107	12.47**
+Sex	.21**	.040	6.30*	.04	.000	.01
+SES (dummy)		.048	2.62†		.029	1.12
SES II	-.03			.04		
SES III	-.24*			-.09		
SES IV-V	-.03			.02		
+Educational Verification	-.12	.014	2.28	-.32**	.094	12.29**
+Ethnicity	-.04	.001	.19	-.08	.005	.66

[a]β in the final model

***$p < .0005$, **$p \leq .01$, *$p \leq .05$, †$p \leq .10$

entered in the model, the total variance explained was 26.0%, R squared change = .074, *F* change (3, 99) = 3.31, *p* = .023. Sex, educational verification, and ethnicity variables did not add significantly to the model, *p*s > .10. The overall model was a significant predictor of MEAP English Language Arts scores, *F* (8, 97) = 4.31, *p* < .0005. In the final model, reading grade, CTAS-Total score, SES-II, and SES-III variables were significant predictors of English Language Arts MEAP scores, with reading grade recording the highest beta value (β = .31, *p* = .001) in comparison to the other variables, SES-II (β = -.30, *p* = .008), SES-III (β = -.27, *p* = .011), and CTAS-Total (β = -.20, *p* = .038).

The regression analysis of fourth grade students' Math MEAP scores as the dependent variable indicated that math grade explained 6.5% of the variance in fourth graders' Math MEAP scores, *F* (1, 104) = 7.19, *p* = .009. After the CTAS-Total was entered in the model, the total variance explained was 9.7%, R squared change = .032, *F* change (1, 103) = 3.65, *p* = .059. Sex, SES, educational verification, and ethnicity variables did not add significantly to the model, *p*s > .10. The overall model was a significant predictor of MEAP Math scores, *F* (8, 97) = 2.28, *p* = .028. In the final model, math grade and SES-III variables were significant predictors of Math MEAP scores, with SES-III recording the highest beta value (β = -.24, *p* = .039) in comparison to math grade (β = .23, *p* = .028). Table 8 summarizes how predictive students' CTAS-Total scale, academic grades, and demographic characteristics were in accounting for variance in English Language Arts and Math MEAP scores for students in grade four.

Grade 5 The regression analysis of fifth grade students' English Language Arts MEAP scores as the dependent variable indicated that reading grade explained 19.0% of

Table 8

Multiple Regression Of Predictors of MEAP Scores in Grade 4

	MEAP					
	English Language Arts			Math		
Additive Model	β^a	ΔR^2	F-Δ	β^a	ΔR^2	F-Δ
+Achievement		.141	17.09**		.065	7.19**
Reading Grade	.31**					
Math Grade				.23*		
+CTAS-Total	-.20*	.037	4.67*	-.15	.032	3.65†
+Sex	.15	.007	.90	-.06	.008	.96
+SES (dummy)		.074	3.31*		.049	1.90
SES II	-.30**			-.20†		
SES III	-.27*			-.24*		
SES IV-V	-.13			-.07		
+Educational Verification	-.03	.001	.10	-.04	.001	.12
+Ethnicity	.04	.002	.22	-.06	.003	.38

[a] β in the final model

*** $p < .0005$, ** $p \leq .01$, * $p \leq .05$, † $p \leq .10$

the variance in fifth graders' English Language Arts MEAP scores, F (1, 78) = 18.26, $p <$ .0005. When the CTAS-Total was entered in the model, the total variance explained was 22.6%, R squared change = .036, F change (1, 77) = 3.62, p = .061. After SES was entered in the model, the total variance explained was 34.5%, R squared change = .102, F change (3, 73) = 3.80, p = .014. After educational verification was entered in the model, the total variance explained was 37.5%, R squared change = .029, F change (1, 72) = 3.40, p = .07. Sex and ethnicity variables did not add significantly to the model, $ps > .10$. The overall model was a significant predictor of MEAP English Language Arts scores, F (8, 71) = 5.31, $p <$ .0005. In the final model, student reading grade and SES-III variables were significant predictors of English Language Arts MEAP scores, with reading grade recording the highest beta value (β = .36, p = .001) in comparison to SES-III (β = -.35, p = .007).

The regression analysis of fifth grade students' Math MEAP scores as the dependent variable indicated that math grade explained 3.2% of the variance in fifth graders' Math MEAP scores, F (1, 78) = 2.60, p = .11. After the CTAS-Total was entered in the model, the total variance explained was 11.2%, R squared change = .080, F change (1, 77) = 6.91, p = .01. After sex was entered in the model, the total variance explained was 18.1%, R squared change = .069, F change (1, 76) = 6.42, p = .01. After educational verification was entered in the model, the total variance explained was 25.8%, R squared change = .041, F change (1, 72) = 4.02, p = .05. SES and ethnicity variables did not add significantly to the model, $ps > .10$. The overall model was a significant predictor of MEAP Math scores, F (8, 71) = 3.10, p = .005. In the final model, sex and educational verification were the only significant predictors of Math MEAP scores, with sex (β = -

.33, p = .004) recording a higher beta value than educational verification (β = -.21, p = .05).

The regression analysis of fifth grade students' Science MEAP scores as the dependent variable indicated that science grade explained 13.9% of the variance in fifth graders' Science MEAP scores, F (1, 78) = 12.55, p = .001. When CTAS-Total was entered in the model, the total variance explained was 15.0%, R squared change = .012, F change (1, 77) = 1.07, p = .30, indicating that CTAS-Total did not add significantly to the model. After sex was entered in the model, the total variance explained was 22.4%, R squared change = .074, F change (1, 76) = 7.22, p = .009. SES, educational verification, and ethnicity variables did not add significantly to the model, ps > .10. The overall model was a significant predictor of MEAP Science scores, F (8, 71) = 3.48, p = .002. In the final model, science grade and sex variables were significant predictors of Science MEAP scores, with science grade recording a higher beta value (β = .43, p = .001) than sex (β = -.27, p = .02). Tables 9 summarizes how predictive students' CTAS-Total scale, academic grades, and demographic characteristics were in accounting for variance in English Language Arts, Math, and Science MEAP scores for students in grade five.

Question Four

Aim 1

Mann-Whitney U tests and Kruskal-Wallis tests were used to examine differences in test anxiety resulting from differences in student demographic variables, including sex, ethnicity, educational verification, socioeconomic status, and grade. Analyses were conducted separately to examine differences in test anxiety scores in relation to the MEAP and classroom testing. Table 10 summarizes differences in students' test anxiety

Table 9

Multiple Regression Of Predictors of MEAP Scores in Grade 5

	MEAP								
	English Language Arts			Math			Science		
Additive Model	β^a	ΔR^2	F-Δ	β^a	ΔR^2	F-Δ	β^a	ΔR^2	F-Δ
+Achievement		.190	18.26***		.032	2.60		.139	12.55**
Reading Grade	.36**								
Math Grade				.18†					
Science Grade							.43**		
+ CTAS-Total	-.19†	.036	3.62†	-.18†	.080	6.91**	-.04	.012	1.07
+Sex	-.141	.017	1.66	-.33**	.069	6.42*	-.27**	.074	7.22**
+SES (dummy)		.102	3.80*		.035	1.09		.042	1.39
SES II	-.02			-.09			.13		
SES III	-.35**			-.17			-.08		
SES IV-V	-.06			-.15			-.11		
+Educational Verification	-.18†	.029	3.40†	-.21*	.041	4.02*	.13	.016	1.57*
+Ethnicity	-.00	.000	.00	-.04	.001	.14	-.01	.000	.01

[a]β in the final model; *** $p < .0005$, ** $p \leq .01$, * $p \leq .05$, † $p \leq .10$

Table 10

Differences in Test Anxiety in Relation to the MEAP Testing Condition by Demographic Variables[a]

Measure	CTAS Total			BASC-2-TA		
	n	*z*	*p*[b]	*n*	*z*	*p*[b]
Sex	301	-4.35	<.0005	305	-4.09	<.0005
Ethnicity	302	-.60	.55	306	-1.18	.24
Educational Verification	295	-.73	.46	299	-.91	.36
	n	χ^2	*p*[b]	*n*	χ^2	*p*[b]
Grade	305	6.90	.03	309	11.46	.003
SES	282	2.23	.69	286	2.69	.61

[a]MEAP: Michigan Educational Assessment Program; CTAS: Children's Test Anxiety Scale; BASC-2-TA: Behavioral Assessment Scale for Children, Second Edition, Test Anxiety Subscale; SES: Hollingshead Socioeconomic Codes: 1: High Class, 2: Middle Class, 3: Working Class, 4 and 5: Low Class

[b]Bonferroni correction for multiple comparisons: $p \leq .01$

scores on the MEAP due to demographic characteristic differences

Differences in students' test anxiety scores due to student demographic characteristics on the CTAS-Total scale in relation to the MEAP were examined first. A Mann-Whitney *U* test revealed a significant difference in CTAS-Total scores of males (*Md* = 48, *n* = 140) and females (*Md* = 56, *n* = 161), *U* = 7998.50, z = -4.35, *p* < .0005. The effect size of this difference was small, r = -.25. There was no significant difference in the CTAS-Total scores of Caucasians (*Md* = 53, *n* = 250) and Non-Caucasians (*Md* = 50.50, *n* = 52), *U* = 6154.50, z = -.60, *p* = .55. There was no significant difference in the CTAS-Total scores of general education students (*Md* = 53, *n* = 268) and special education students (*Md* = 49, *n* = 27), *U* = 3308.50, z = -.73, *p* = .46. A Kruskal-Wallis Test revealed a marginally significant difference in CTAS-Total scores across three grade levels (Gp1, *n* = 102: Grade 3, Gp2, *n* = 115: Grade 4, Gp3, *n* = 88: Grade 5), χ^2 (2, *n* = 305) = 6.90, *p* = .03. Mann-Whitney *U* tests between the three grade levels revealed a significant difference in the CTAS-Total scores of grade 3 students (*Md* = 50.50, *n* = 102) and grade 4 students (*Md* = 56 *n* = 115), *U* = 4921,50, z = -2.05, *p* = .04, and between grade 4 students (*Md* = 56, *n* = 115) and grade 5 students (*Md* = 49, *n* = 88), *U* = 4069.50, z = -2.39, *p* = .02. There was no significant difference between reported test anxiety between third and fifth grade students. A Kruskal-Wallis Test also revealed no significant difference in CTAS-Total scores across the five socioeconomic groups (Gp1, *n* = 65: SES I, Gp2, *n* = 144: SES II, Gp3, *n* = 57: SES III, Gp4, *n* = 13: SES IV, Gp5, *n* = 3: SES V), χ^2 (4, *n* = 282) = 2.23, *p* = .69.

Differences in students' test anxiety scores due to demographic characteristics on the BASC-2-TA scale in relation to the MEAP were examined second. A Mann-Whitney

U test revealed a significant difference in the BASC-2-TA scores of males (*Md* = 3, *n* = 144) and females (*Md* = 5, *n* = 161), *U* = 8460, *z* = -4.09, *p* < .0005. The effect size of this difference was small, *r* = -.23. There was no significant difference in the BASC-2-TA scores of Caucasians (*Md* = 4, *n* = 254) and non-Caucasians (*Md* = 3, *n* = 52), *U* = 5923.00, *z* = -1.18, *p* = .24. There was no significant difference in the BASC-2-TA scores of general education students (*Md* = 4, *n* = 270) and special education students (*Md* = 3, *n* = 29), *U* = 3515.50, *z* = -.91, *p* = .36. A Kruskal-Wallis Test revealed a significant difference in BASC-2-TA scores across students at the three grade levels (Gp1, *n* = 103: Grade 3, Gp2, *n* = 117: Grade 4, Gp3, *n* = 89: Grade 5), χ^2 (2, *n* = 309) = 11.46, *p* = .003. Mann-Whitney *U* tests between the three grade levels revealed a significant difference in the BASC-2-TA scores of grade 3 students (*Md* = 3, *n* = 103) and grade 4 students (*Md* = 5 *n* = 117), *U* = 4470.5, z = -3.32, *p* = .001. There was no significant difference in reported test anxiety between fourth and fifth grade or third and fifth grade students. There was also no significant difference in BASC-2-TA scores across the five socioeconomic groups (Gp1, *n* = 65: SES I, Gp2, *n* = 147: SES II, Gp3, *n* = 57: SES III, Gp4, *n* = 14: SES IV, Gp5, *n* = 3: SES V), χ^2 (4, *n* = 286) = 2.69, *p* = .61.

Next, differences in students' test anxiety scores in relation to classroom testing were examined. Table 11 summarizes the differences in students' test anxiety scores in relation to classroom testing due to demographic characteristics. Differences in students' test anxiety scores on the CTAS-Total scale in relation classroom testing were examined first. A Mann-Whitney *U* test revealed a significant difference in the CTAS-Total scores of males (*Md* = 44, *n* = 144) and females (*Md* = 51 *n* = 171), *U* = 9382, z = -3.64, *p* <

Table 11

Differences in Test Anxiety in Relation to the Classroom Testing Condition by Demographic Variables

	CTAS-Total[a]			BASC-2-TA			TASC		
Measure	*n*	*z*	*p*[b]	*n*	*z*	*p*[b]	*n*	*z*	*p*[b]
Sex	315	-3.64	<.0005	319	-4.59	<.0005	318	-4.73	<.0005
Ethnicity	317	-.33	.74	321	-.34	.73	320	-.53	.60
Educational Verification	308	-.98	.33	312	-.37	.71	311	-.47	.64
	n	χ^2	*p*[b]	*n*	χ^2	*p*[b]	*n*	χ^2	*p*[b]
Grade	318	2.73	.26	322	1.01	.60	321	4.93	.09
SES	293	7.22	.13	297	2.18	.70	296	2.20	.70

[a]CTAS: Children's Test Anxiety Scale; BASC-2-TA: Behavioral Assessment Scale for Children, Second Edition, Test Anxiety Subscale; TASC: Test Anxiety Scale for Children; SES: Hollingshead Socioeconomic Codes: 1: High Class, 2: Middle Class, 3: Working Class, 4 and 5: Low Class.

[b]Bonferroni correction for multiple comparisons: $p \leq .01$

.0005. The effect size of this difference was small, $r = -.21$. There was no significant difference in the CTAS-Total scores of Caucasians ($Md = 47$, $n = 263$) and non-Caucasians ($Md = 46.50$, $n = 54$), $U = 6898.50$, $z = -.33$, $p = .74$. There was no significant difference in the CTAS-Total scores of general education students ($Md = 48$, $n = 278$) and special education students ($Md = 44.50$, $n = 30$), $U = 3718$, $z = -.7398$ $p = .33$. A Kruskal-Wallis Test revealed no significant difference in CTAS-Total scores across the three grade levels (Gp1, $n = 115$: Grade 3, Gp2, $n = 115$: Grade 4, Gp3, $n = 88$: Grade 5), χ^2 (2, $n = 318$) = 2.73, $p = .26$. There was also no significant difference in CTAS-Total scores in relation to classroom testing across the five socioeconomic groups (Gp1, $n = 70$: SES I, Gp2, $n = 152$: SES II, Gp3, $n = 54$: SES III, Gp4, $n = 12$: SES IV, Gp5, $n = 5$: SES V), χ^2 (4, $n = 293$) = 7.22, $p = .13$.

Differences in students' test anxiety scores on BASC-2-TA scale in relation classroom testing were examined next. A Mann-Whitney U test revealed a significant difference in the BASC-2-TA scores of males ($Md = 2$, $n = 146$) and females ($Md = 5$, $n = 173$), $U = 8883$, $z = -4.59$, $p < .0005$. The effect size of this difference was small, $r = -.26$. There was no significant difference in the BASC-2-TA scores of Caucasians ($Md = 4$, $n = 267$) and non-Caucasians ($Md = 3$, $n = 54$), $U = 6996.00$, $z = -.34$, $p = .73$. There was no significant difference in the BASC-2-TA scores of general education students ($Md = 4$, $n = 281$) and special education students ($Md = 3$, $n = 31$), $U = 4180.50$, $z = -.37$, $p = .71$. A Kruskal-Wallis Test revealed no significant difference in BASC-2-TA scores across the three grade levels (Gp1, $n = 115$: Grade 3, Gp2, $n = 117$: Grade 4, Gp3, $n = 90$: Grade 5), χ^2 (2, $n = 322$) = 1.01, $p = .60$. There was also no significant difference in BASC-2-TA scores across the five socioeconomic groups (Gp1, $n = 70$: SES I, Gp2, $n =$

154: SES II, Gp3, n = 55: SES III, Gp4, n = 13: SES IV, Gp5, n = 5: SES V), χ^2 (4, n = 297) = 2.18, p = .70.

Differences in students' test anxiety scores on the TASC scale in relation classroom testing were examined last. A Mann-Whitney U test revealed a significant difference in the TASC scores of males (Md = 5, n = 147) and females (Md = 10, n = 171), U = 8710.50, z = -4.73, p < .0005. The effect size of this difference was small, r = -.27. There was no significant difference in the TASC scores of Caucasians (Md = 7, n = 264) and non-Caucasians (Md = 7, n = 56), U = 7061.00, z = -.53, p = .60. There was no significant difference in the TASC scores of general education students (Md = 8, n = 279) and special education students (Md = 7, n = 32), U = 4236, z = -.47, p = .64. A Kruskal-Wallis Test revealed no significant difference in TASC scores across the three grade levels (Gp1, n = 116: Grade 3, Gp2, n = 116: Grade 4, Gp3, n = 89: Grade 5), χ^2 (2, n = 321) = 4.93, p = .09. There was also no significant difference in TASC scores across the five socioeconomic groups (Gp1, n = 70: SES I, Gp2, n = 153: SES II, Gp3, n = 55: SES III, Gp4, n = 13: SES IV, Gp5, n = 5: SES V), χ^2 (4, n = 296) = 2.20, p = .70.

Aim 2

Hierarchical multiple regression analyses were used to assess the ability of demographic variables, including student grade, sex, SES, educational verification, and ethnicity, to predict scores on the different test anxiety measures. Data transformations of test anxiety scores were completed to account for the non-normal distribution of test anxiety scores. Following logarithmic data transformations of CTAS-Total variable and square root data transformations of BASC-2-TA and TASC variables (Pallant, 2007), the test anxiety variables were found to more closely meet the assumption of normality.

Preliminary analyses were conducted for each regression analysis to ensure no violation of the assumptions of multiple regression. Tolerance and variance inflation factor scores (VIFs) were examined and no indications of multicollinearity were identified (Tols. > .10 and VIFs < 10). Normal probability plots of the regression standardized residuals and scatterplots of the standardized residuals were examined and no violations of normality, linearity, homoscedasticity, or independence of residuals were identified for any of the regression analyses. Table 12 summarizes how predictive demographic variables were in accounting for variance in test anxiety scores by each test anxiety measure across both testing conditions. In each regression analysis, sex was entered in Step 1 (female = 1, male = 0), grade was dummy coded in Step 2: Grade4 (Grade 4 = 1, other = 0) and Grade5 (Grade 5 = 1, other = 0), ethnicity was entered in Step 3 (non-Caucasian = 1, Caucasian = 0), educational verification was entered in Step 4 (special education = 1, general education = 0), and socioeconomic status was dummy coded in Step 5: SES II (SES-II = 1, other = 0), SES-III (SES-III = 1, other = 0), and SES-IV-V (SES-IV-IV = 1, other = 0).

The value of demographic variables in predicting test anxiety scores in relation to the MEAP were considered first. The first regression analysis of CTAS-Total as the dependent variable indicated that sex explained 6.2% of the variance in CTAS-Total scores, $F(1, 280) = 18.40$, $p < .0005$. After grade was entered in the model, the total variance explained was 8.6%, R squared change = .024, F change $(2, 278) = 3.63$, $p = .03$. Student ethnicity, educational verification, and SES did not add significantly to the model. The final model was a significant predictor of CTAS-Total scores in relation to

Table 12

Multiple Regression of Demographic Variables Predicting Test Anxiety Scores

	MEAP Assessment						Classroom Tests								
	CTAS-Total[b]			BASC-2-TA[b]			CTAS-Total[b]			BASC-2-TA[b]			TASC[b]		
Additive Model	β^a	ΔR^2	F-Δ	β^a	ΔR^2	F-Δ	β^a	ΔR^2	F-Δ	β^a	ΔR^2	F-Δ	β^a	ΔR^2	F-Δ
Sex	.24**	.06	18.40**	.24**	.05	15.51**	.19*	.04	12.29*	.26**	.07	20.74**	.26**	.07	22.10**
+Grade (dummy)		.02	3.63†		.03	4.77*		.01	1.55		.00	.61		.01	2.28
Grade 4	.16†			.20**			.11			.08			.08		
Grade 5	-.02			.11			-.02			.01			-.08		
+ Ethnicity	-.07	.00	.57	-.10	.01	1.61	-.02	.00	.04	.01	.00	.08	.02	.00	.15
+ Educational Verification	-.00	.00	.03	-.01	.00	.00	-.04	.00	.20	.03	.00	.31	.01	.00	.07
+ SES (dummy)		.01	1.28		.02	1.69		.04	3.60*		.01	.74		.01	1.25
SES II	.02			-.03			.01			.01			-.07		
SES III	.09			-.07			.19*			.09			.06		
SES IV-V	.10			.10			.09			.04			.01		

[a]β in the final model; [b]Transformed variables; $^{**}p < .0005$, $^{*}p \leq .01$, $^{\dagger}p < .05$

the MEAP, F (8, 273) = 3.80, p < .0005. In the final model, sex and grade were the only significant predictors of CTAS-Total scores on the MEAP, with sex recording a higher beta value (β = .24, p < .0005) than Grade4 (β = .16, p = .02) or Grade5 (β = -.02, p = .74) variables.

The second regression analysis of BASC-2-TA as the dependent variable indicated that sex explained 5.2% of the variance in BASC-2-TA scores in relation to the MEAP, F (1, 284) = 15.51, p < .0005. After grade was entered in the model, the total variance explained was 8.3%, R squared change = .031, F change (2, 278) = 4.77, p = .009. Student ethnicity, educational verification, and SES did not add significantly to the model. The final model was a significant predictor of BASC-2-TA scores in relation to the MEAP, F (8, 277) = 4.04, p < .0005. In the final model, sex and grade were the only significant predictors of BASC-2-TA scores in relation to the MEAP, with sex recording a higher beta value (β = .24, p < .0005) than Grade4 (β = .20, p = .002) or Grade5 (β = .11, p = .08) variables.

Next, three regression analyses examining the value of demographic variables in predicting test anxiety measure scores in relation to classroom testing were conducted using the same five-step model. The first regression analysis of CTAS-Total as the dependent variable indicated that sex explained 4.1% of the variance in CTAS-Total scores in relation to classroom testing, F (1, 291) = 12.29, p = .001. Student grade, ethnicity, and educational verification did not add significantly to the model. After SES was entered as the final predictor in the model, the total variance explained was 8.6%, R squared change = .035, F change (3, 284) = 3.60, p = .014. The final model was a significant predictor of CTAS-Total scores in relation to classroom testing, F (8, 284) =

3.52, $p = .001$. In the final model, sex and SES were the only significant predictors of CTAS-Total scores in relation to classroom testing, with sex recording a higher beta value ($\beta = .19, p = .001$) than the SES-II ($\beta = .01, p = .89$), SES-III ($\beta = .19, p = .008$), or SES-IV-V ($\beta = .09, p = .18$) variables.

The second regression analysis of the BASC-2-TA as the dependent variable indicated that sex explained 6.6% of the variance in BASC-2-TA scores in relation to classroom testing, $F(1, 294) = 20.74, p < .0005$. Student grade, ethnicity, educational verification, and SES did not add significantly to the model. The final model was a significant predictor of BASC-2-TA scores in relation to classroom testing, $F(8, 288) = 3.04, p = .003$. In the final model, sex was the only significant predictor of BASC-2-TA scores in relation to classroom testing, beta value ($\beta = .26, p < .0005$).

The third regression analysis of the TASC as the dependent variable indicated that sex explained 7.0% of the variance in TASC scores in relation to classroom testing, $F(1, 294) = 22.10, p < .0005$. Student grade, ethnicity, educational verification, and SES did not add significantly to the model. The final model was a significant predictor of TASC scores in relation to classroom testing, $F(8, 287) = 3.84, p < .0005$. In the final model, sex was the only significant predictor of TASC scores in relation to classroom testing, beta value ($\beta = .26, p < .0005$).

Question Five

Aim 1

The relationship between all measures of test anxiety were investigated using two-tailed Spearman rho correlation coefficient analyses for each testing condition. First, the relationship between the CTAS and the BASC-2-TA scales in relation to the MEAP

was examined and the findings are summarized in Table 13. All measures were significantly related to one another at the $p < .0005$ level. There was a strong, positive correlation between the CTAS-Total scale and BASC-2-TA subscale, $r_s = .71$, $n = 305$, $p < .0005$. Additionally, there were strong, positive correlations between the CTAS-Thoughts and CTAS-Autonomic subscales and the BASC-2-TA subscale ($r_s = .72$ and $r_s = .64$, respectively). The CTAS Off-task subscale was only moderately related to the BASC-2-TA subscale ($r_s = .41$).

As expected, while the three subscales of the CTAS were significantly correlated with one another, correlations between subscales were less strong than the strength of the relationship between the subscales and the CTAS-Total scale as the subscales were designed to measure different types of test anxiety symptoms. There was a moderate, positive correlation between the CTAS-Thought and CTAS-Off-task subscales, $r_s = .45$, $n = 306$, $p < .0005$. There was a moderate, positive correlation between the CTAS-Autonomic and CTAS-Off-task subscales, $r_s = .46$, $n = 305$, $p < .0005$. There was a strong, positive relationship between the CTAS-Thought and CTAS-Autonomic subscales, $r_s = .70$, $n = 306$, $p < .0005$.

Next, the relationship between the CTAS, BASC-2-TA, and TASC scales in relation to classroom testing was examined and the findings are summarized in Table 14. All measures were significantly related to one another at the $p < .0005$ level. There was a strong, positive correlation between the CTAS-Total scale and BASC-2-TA scale, $r_s = .73$, $n = 317$, $p < .0005$. There was also a strong, positive correlation between the CTAS-Total scale and TASC scale, $r_s = .77$, $n = 316$, $p < .0005$. Finally, there was a strong,

Table 13

Correlations between Measures of Test Anxiety in Relation to the MEAP[a]

Measure	1	2	3	4	5
1. CTAS-Total	1	.90**	.73**	.84**	.71**
2. CTAS-Thoughts		1	.45**	.70**	.72**
3. CTAS-Off-task			1	.46**	.41**
4. CTAS-Autonomic				1	.64**
5. BASC-2-TA					1

[a]MEAP: Michigan Educational Assessment Program; CTAS: Children's Test Anxiety Scale; BASC-2-TA: Behavioral Assessment Scale for Children, Second Edition, Test Anxiety Subscale

** $p < .0005$

Table 14

Correlations between Measures of Test Anxiety in Relation to Classroom Testing[a]

Measure	1	2	3	4	5	6
1. CTAS-Total	1	.90**	.81**	.84**	.73**	.77**
2. CTAS-Thoughts		1	.55**	.70**	.70**	.74**
3. CTAS-Off-task			1	.58**	.49**	.50**
4. CTAS-Autonomic				1	.69**	.74**
5. BASC-2-TA					1	.78**
6. TASC						1

[a]MEAP: Michigan Educational Assessment Program; CTAS: Children's Test Anxiety Scale; BASC-2-TA: Behavioral Assessment Scale for Children, Second Edition, Test Anxiety Subscale; TASC: Test Anxiety Scale for Children

** $p < .0005$

positive correlation between the BASC-2-TA subscale and the TASC scale, $r_s = .78$, $n = 318$, $p < .0005$. Additionally, there were strong positive correlations between the CTAS-Thoughts and CTAS-Autonomic subscales and the BASC-2-TA scale ($r_s = .70$ and $r_s = .69$, respectively). There were also strong positive correlations between the CTAS-Thoughts and CTAS-Autonomic subscales and the TASC scale ($r_s = .74$ and $r_s = .74$, respectively). There was only a moderate relationship between the CTAS-Off-task subscale and the BASC-2-TA and TASC scales ($r_s = .49$ and $r_s = .50$, respectively).

As expected, while the three subscales of the CTAS were significantly correlated with one another, correlations between subscales were less strong than the strength of the relationship between the subscales and the CTAS-Total scale as the subscales were designed to measure different types of test anxiety symptoms. There was a moderate, positive correlation between the CTAS-Thought and CTAS-Off-task subscales, $r_s = .55$, $n = 319$, $p < .0005$. There was a moderate, positive correlation between the CTAS-Autonomic and CTAS-Off-task subscales, $r_s = .58$, $n = 321$, $p < .0005$. There was a strong, positive relationship between the CTAS-Thought and CTAS-Autonomic subscales, $r_s = .70$, $n = 318$, $p < .0005$.

Finally, a Kappa Measure of Agreement was calculated for children classified as high test-anxious versus moderately or low test-anxious using the TASC (Definition 2) and the CTAS-Total (Definition 1) classification systems. The Kappa Measure of agreement indicated moderate agreement between the two classification systems, $\kappa = .45$, $p < .0005$. Out of the 65 cases classified as high test-anxious by the TASC, 26 were also classified as high test-anxious on the CTAS indicating a sensitivity of 40%.

Aim 2

The relationship between test anxiety in relation to classroom testing and the MEAP was investigated using two-tailed Spearman rho correlation analyses for the two test anxiety measures repeated across conditions. These findings are summarized in Table 15. As hypothesized, all measures were significantly related to one another across testing conditions at the $p < .0005$ level. The strongest relationship between reported test anxiety across the MEAP and classroom testing conditions was found for the CTAS-Total scale, $r_s = .73$, $n = 292$, $p < .0005$. Students' scores on the CTAS-Total scale in relation to classroom testing explained 53% of the variance in students' scores on the CTAS-Total scale in relation to the MEAP. The correlations between the three subscales of the CTAS in relation to classroom testing and the MEAP ranged from $r_s = .66$ to .69. The weakest relationship between reported test anxiety across the MEAP and classroom testing conditions was found for the BASC-2-TA subscale, $r_s = .61$, $n = 298$, $p < .0005$. Students' scores on the BASC-2-TA subscale in relation to classroom testing explained only 37% of the variance in students' scores on the BASC-2-TA subscale in relation to the MEAP.

Table 15

Correlation between Measures Assessing Test Anxiety in Relation to the MEAP and Classroom Testing[a]

Measure	*n*	r_s	*p*
CTAS-Total	292	.73	<.0005
CTAS-Thoughts	294	.66	<.0005
CTAS-Off-task	295	.69	<.0005
CTAS-Autonomic	295	.67	<.0005
BASC-2-TA	298	.61	<.0005

[a]MEAP: Michigan Educational Assessment Program; CTAS: Children's Test Anxiety Scale; BASC-TA: Behavioral Assessment Scale for Children, Second Edition, Test Anxiety Subscale

CHAPTER 5

DISCUSSION

The purpose of this study was to critically examine how children in elementary school perceive high-stakes and classroom testing situations in order to enhance our limited understanding of the impact of test anxiety in high-stakes testing situations and to advance our understanding of test anxiety in young children in general. To this end, this study examined differences in test anxiety reported by children in grades three through five in relation to classroom testing and high-stakes testing. Additionally, teacher perceptions of student anxiety, teacher anxiety, and student performance were investigated. This aspect of the study enhances our understanding of test anxiety by enabling a comparison between teachers' perceptions of student test anxiety and students' self-reported test anxiety. Further, this study examined differences in how children were classified as low, moderate, or high test-anxious using three different classification systems. Next, differences in student performance on the high-stakes assessment were examined between students classified as being low, moderate, or high test-anxious.

In addition, the ability of student demographic characteristics to predict student test anxiety was examined. Similarly, the ability of student grades, test anxiety, and demographic characteristics to predict high-stakes test performance was examined. Finally, the relationship between the three scales used to measure test anxiety was examined. The current study offers multiple unique contributions to the literature by 1) by directly examining high-stakes test anxiety among young school children, 2) examining differences in test anxiety by test condition, 3) examining differences in test

performance by test anxiety level, and 4) by examining differences in test anxiety classifications across different measurement tools.

Student Self-Report of Test Anxiety

This study examined differences between students' self-report of test anxiety in relation to high-stakes and classroom testing situations. Previous test anxiety literature suggests that test anxiety is associated with impaired test performance and impaired knowledge acquisition in academic skill areas (Sarason et al., 1960; Zeidner, 1998). In the current study, students as a whole reported significantly more test anxiety in relation to the MEAP assessment than in relation to classroom testing across both measures of test anxiety repeated for the two testing conditions. There were small, but significant differences between students' self-report of test anxiety in relation to the MEAP and classroom testing for both the CTAS-Total scale and the BASC-2-TA scale, $r = -.21$ and $r = -.10$, respectively. These results are consistent with the hypothesis that as a whole, students perceive high-stakes testing situations as more stressful and anxiety-provoking than typical testing situations that occur as a regular part of the curriculum. Similarly, students reported significantly more cognitive and physiological symptoms of test anxiety in relation to the MEAP than in relation to classroom testing. When considered within the context of previous research that suggests worry is more strongly related to impaired test performance than emotionality in both laboratory and applied-testing situations (Hembree, 1988, Holroyd et al., 1978, Liebert & Morris, 1967), this finding indicates that the significantly greater cognitive and physiological test anxiety symptoms reported by students in relation to the high-stakes assessment may be associated with performance impairments.

The finding that on the whole young children experienced greater test anxiety in the high-stakes testing condition as opposed to the classroom testing condition is first of its kind in the literature. No previous study has directly compared test anxiety across both high-stakes and typical testing conditions. The only study that offers some data for comparison is a study comparing a mock high-stakes examination taken as part of test preparation for an actual high-stakes terminal high-school examination (Putwain, 2008). However, this study differs significantly from the current study in that students taking the mock examination were taking the practice "high-stakes" examination for the first time. In contrast to the current study findings, Putwain found that students taking the mock or low-stakes examination experienced greater test anxiety than students taking the high-stakes terminal examination. One possible explanation for the difference in the findings may be that students in the current study had a depth of experience with classroom testing situations in comparison to the relatively novel experience of the high-stakes assessment whereas the mock examination in the Putwain study was a completely novel experience for the participants. Alternatively, students taking the mock examination in the Putwain study may have perceived the practice test as having significant stakes associated with it. While it is not possible to identify a causal mechanism for the differences in reported test anxiety in either the current study or the Putwain study, the finding in the current study that as a whole, young children reported significantly more test anxiety about the high-stakes condition than the classroom testing condition suggests that students perceived the two testing conditions significantly differently.

Interestingly, students' average test anxiety scores across both the high-stakes testing condition and the classroom testing condition fell within one standard deviation of

the normative samples for the BASC-2-TA and the CTAS (Reynolds & Kamphaus, 2004; Wren, personal communication, March 14, 2008). This finding would suggest that although there were small and significant increases in test anxiety in relation to high-stakes testing, these differences did not result in clinically significant increases in test anxiety for the *overall* sample. Similarly, students' average test anxiety scores on the cognitive and physiological subscales of the CTAS also fell within one standard deviation of the normative samples (Wren), suggesting that there was no clinically significant increase in cognitive and physiological symptoms of test anxiety for the *overall* sample. In light of the significant difference in overall test anxiety as well as in cognitive and physiological aspects of test anxiety across the two testing conditions, differences in rates of low, moderate, and high test anxiety among students are examined following discussion of teachers' perceptions of test anxiety among their students. As discussed above, if the increase in reported test anxiety results in differences in rates of students who are classified as low, moderate, and high test-anxious across test conditions, this increase would be significant given the established relationship between test anxiety and impaired task performance.

Teacher Perceptions of Testing

In addition to examining student self-reported test anxiety, this study examined teacher perceptions of student test anxiety, enabling direct comparisons to be made between teachers' perceptions of student test anxiety and students' self-reported test anxiety. This aspect of the current study adds to the existing literature on teacher perceptions of test anxiety, which is limited by a lack of direct student perception data. In the current study, teachers reported that they believed students experienced significantly

more anticipatory anxiety regarding the high-stakes assessment than classroom testing. This finding is consistent with multiple studies examining teachers' perceptions of the impact of high-stakes testing on children which indicate teachers report heightened anxiety, stress, pressure, and worry among students due to high-stakes testing programs (Abrams et al., 2003; Barksdale-Ladd & Thomas, 2000; Jones & Egley, 2004; Jones et al., 1999). In the current study, this finding is also consistent with the significantly greater test anxiety reported by students in relation to the high-stakes testing condition versus the classroom testing condition. Hence, this study indicates that teachers' perceptions of students' response to the different testing conditions were accurate.

In addition, teachers reported that they had significantly more anxiety about how well students would perform on the high-stakes assessment than classroom testing. Again, this finding is consistent with multiple studies examining teachers' psychological response to high-stakes testing programs which indicate that teachers experience increased anxiety, stress, and pressure due to high-stakes testing programs and that teachers change their instructional patterns to focus on test preparation as a result (Abrams et al., 2003; Barksdale-Ladd & Thomas, 2000; Jones & Egley, 2004; Jones et al., 1999). The teacher questionnaire also indicated that there was a trend toward teachers believing that students' performance on the high-stakes assessment was less representative of their true ability than their performance on classroom tests. Future studies examining whether or not teachers' anxious response to testing affects student self-reported test anxiety or student performance are warranted.

Test Anxiety Classifications and Rates across Measures

This study also examined differences in test anxiety classifications and prevalence

rates of students identified as low, moderate, and high test anxious in the high-stakes testing and classroom testing conditions. Within testing conditions, there were significant differences in how the different measures classified students' test anxiety levels. The largest difference in classification was found between the CTAS-Total and the TASC classification systems in relation to the classroom testing condition. More students were identified as low (53% versus 45%) and high (21% versus 11%) test anxious and fewer children identified as moderate (26% versus 44%) test anxious using the TASC classification system than the CTAS-Total classification system, respectively.

This finding demonstrates that test anxiety classifications for the same students differed significantly depending upon the measure used to assess test anxiety. Using students' CTAS-Total score to determine test anxiety levels resulted in 11% of students being classified as high test-anxious in relation to classroom tests, whereas 21% of students were classified as high test-anxious using students' TASC scores. This finding offers some explanation for how different researchers have identified such disparate rates of test anxiety in previous research. Previous estimates of high test anxiety rates have varied widely based on the different methods researchers have used to define clinically significant levels of test anxiety (Zeidner, 1998). For example, King and Ollendick (1989) reported that the prevalence of test anxiety among school-aged children may range from as little as 10% to as much as 30%, with Hill and Wigfield (1984) estimating that only 10% of children experience impairments in test performance as a result of test anxiety. Alternatively, Turner and colleagues (1993) reported that the rate of test anxiety might be as high as 41% among African American children. Since these rates vary so widely, it is difficult for a researcher determine whether or not these studies are

measuring the same test anxiety construct.

This is the first study that directly examines differences in classification systems for test anxiety. It is clear that prevalence rate estimates of test anxiety cannot be directly compared across studies using different measures because the established classification systems for the different measures result in significantly different classifications for the same subjects. In particular, the TASC classification system established by Beidel and Turner (1988) appears to stratify significantly more students into the extreme high or low ranges of test anxiety, with only 26% of students falling outside of these classifications. Alternatively, 44% of students fell in the moderate, or "average" range using the CTAS-Total to classify student test anxiety on classroom tests. In the current study, the CTAS-Total classification system aligns closely with the suggested 10% prevalence rate of children who experience clinically significant impairment as a result of test anxiety (Hill & Wigfield, 1984; King & Ollendick, 1989).

Similarly, the classification system that used both CTAS-Total and TASC scores to classify students' test anxiety levels was more conservative in identifying children as either high or low test-anxious than either the CTAS-Total or TASC classification systems alone and may underestimate the number of students who report low and high levels test anxiety. In the high-stakes test condition, fewer students were identified as low (28% versus 32%) and high (4% versus 9%) test anxious and more students identified as moderate (67% versus 59%) test-anxious with the combined classification system than CTAS-Total classification system. Similarly, on classroom test condition, fewer students were identified as low (41% versus 45%) and high (9% versus 11%) test-anxious and more children identified as moderate (50% versus 44%) test-anxious with the combined

classification system than the CTAS-Total classification system.

Given the significant differences in how each classification system identified students, the need for a standard classification system for identifying impairing levels of test anxiety is clear. The high test anxiety classification on the TASC was established as valid by examining students' scores on the TASC in relation to clinical diagnoses of anxiety disorders (Beidel & Turner, 1988). Alternatively, the high test anxiety classification on the CTAS-Total measure was established by determining if students' scores were significantly greater than the mean of the normative sample (Wren, personal communication, March 14, 2008). In neither case was the test anxiety classification made based on impaired student test performance.

In light of previous research indicating that test anxiety is associated with impaired task performance (Hembree, 1988; Holroyd et al., 1978; Liebert & Morris, 1967), we propose that using impaired student performance would be an appropriate and theoretically sound criteria for developing a standard classification system for identifying students as having clinically significant test anxiety. Alternately, the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Text Revision (DSM-IV-TR) defines a mental disorder as a clinically significant behavioral or psychological pattern that is associated with distress or disability (American Psychiatric Association, 2000). Using this as the criteria for clinical impairment could result in also classifying a student as clinically test-anxious if they report significant distress in relation to test anxiety symptoms. These two classification systems may be appropriate for distinguishing between students who experience psychological distress and impaired testing performance from those who experience psychological distress but do not experience

performance impairments. Development and use of one or both of these proposed classification systems may facilitate screening and decision-making about which students are at most risk for educational and/or psychological impairment associated with test anxiety.

Test Anxiety Classifications across Conditions

This study also examined differences in how students were classified as low, moderate, and high test-anxious across the high-stakes and classroom testing conditions. This aspect of the study is unique in that it examined how all students responded to testing using low, moderate, and high classifications, rather than only examining differences between low and high test-anxious students. In the current study, when students were classified using their CTAS-Total scores, rates of high test anxiety ranged from 9 to 11 percent across testing conditions, suggesting that there was no significant difference in the rate of children identified as highly test-anxious across testing conditions. However, there were significant differences in the proportions of students classified as having low, moderate, or high test anxiety. Differences in classification proportions resulted from more students reporting moderate test anxiety (59% versus 44%) and fewer students reporting low test anxiety (32% versus 45%), in relation to the high-stakes testing condition than the classroom testing condition. Similarly, when students were classified on the basis of stable scores across the CTAS-Total and TASC scales, differences in classification proportions resulted from students more frequently reporting moderate test anxiety (67% versus 50%) and fewer children reporting low test anxiety (28% versus 41%) in relation to the high-stakes testing condition than the classroom testing condition.

This is the first study to specifically examine shifts in student test anxiety classifications across high-stakes and classroom testing conditions. Figure 4 shows differences in the rates of students identified with the low, moderate, and high test anxiety across testing conditions using the CTAS-Total classification system. Additionally, Figure 4 illustrates shifts in how children were reclassified across conditions by identifying which students reported low, moderate, and high test anxiety in the high-stakes testing condition according to their test anxiety classification in the classroom testing condition. As can be seen in Figure 4, 56 students, or 42% of the students who were classified as low test-anxious during classroom testing situations reported significantly heightened anxiety in high-stakes testing situations and were reclassified as moderately test anxious in the high-stakes testing situation. Only 4 students, or 3% of the students who were classified as low test-anxious during classroom testing were reclassified as highly test-anxious in the high-stakes testing situation. Similarly, while 71% of the students classified as moderately test-anxious in the classroom testing condition remained classified as moderately test-anxious in the high-stakes condition, 14 (11%) were reclassified as high test-anxious and 23 (18%) were reclassified as low test-anxious. Finally, the least stability was found among the students who reported high test-anxiety in the classroom testing condition, with 21 (66%) being reclassified as moderately test-anxious and 1 (3%) being reclassified as low test-anxious.

In sum, although there was some reclassification of students into each of the three test anxiety levels across conditions, the largest shift occurred in the number of students (n = 56) who reported low test anxiety levels in the classroom testing condition and were then reclassified as moderately test-anxious in the high-stakes testing condition. Overall,

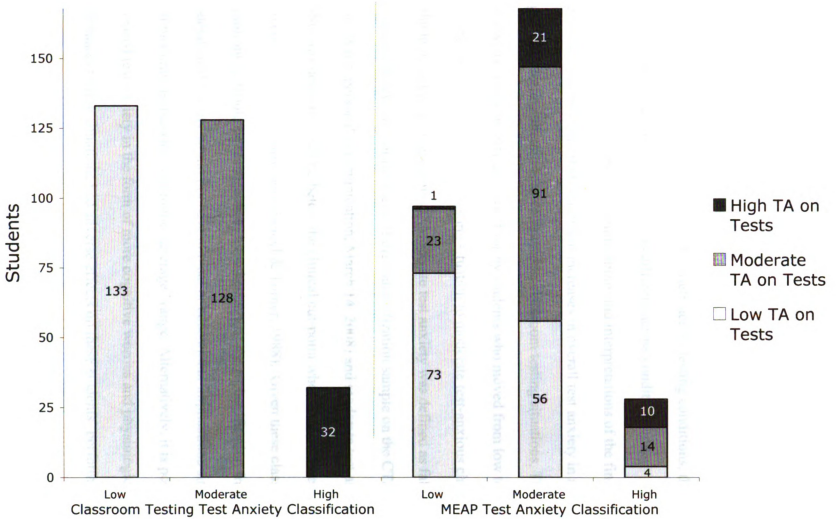


Figure 4: Shifts in Students' Test Anxiety Classifications across Testing Conditions

25% of students reported significant increases in test anxiety symptoms in relation to high-stakes testing and were reclassified into higher test anxiety levels. Of note, however, is the finding that a subset (15%) of students reported decreases in test anxiety in relation to the high-stakes testing condition. It is also interesting that although the same proportion of students reported high test anxiety across testing conditions, the individual students who made up these groups was different across conditions.

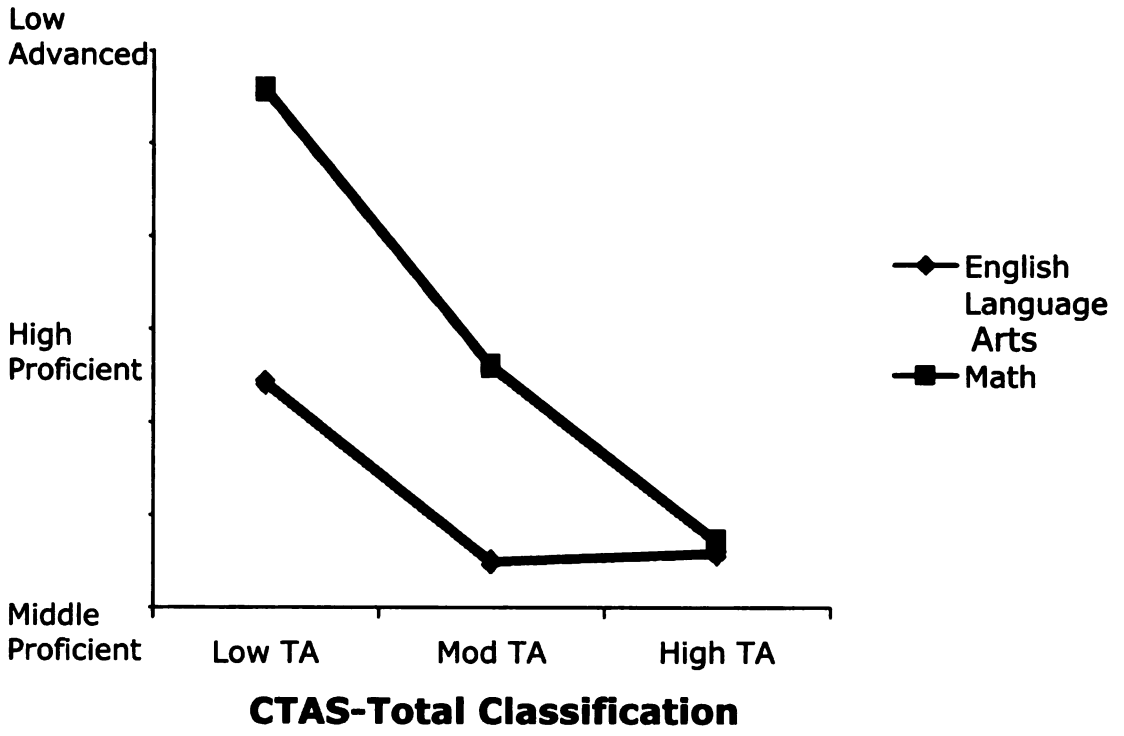
There are several possible implications and interpretations of the finding that on the whole, students' reported significant increases in overall test anxiety in relation to high-stakes testing situations as compared to classroom testing situations. This increase in test anxiety was primarily accounted for by students who moved from low test-anxious classifications in the classroom testing situation to moderate test-anxious classifications in the high-stakes testing situation. Moderate test anxiety was defined as falling within one standard deviation of the mean of the standardization sample on the CTAS-Total scale (Wren, personal communication, March 14, 2008) and moderate test anxiety on the TASC was defined as falling below the clinical cut-point where children were identified with comorbid anxiety disorders (Beidel & Turner, 1988). Given these classification definitions, it would be reasonable to hypothesize that the reclassification from low to moderate test anxiety would not have clinical or educational implications since students with moderate test anxiety fall in the "average" range. Alternatively, it is possible that increased test anxiety in the form of more cognitive worries and physiological arousal in high-stakes testing situations had adverse effects on these students' performance during the testing situation and on students' psychological well-being.

In order to understand the educational implications of students' heightened test

anxiety in relation to high-stakes testing, it is necessary to directly compare students' performance on the high-stakes test with their reported test anxiety. While previous research among elementary school children suggests that high test-anxious children have significantly impaired school achievement and school grades in comparison to low-anxious children (Hill & Sarason, 1966), achievement differences between moderately test-anxious children and low and high test-anxious children have not been previously studied. A unique aspect of the current study is the ability to examine the relationship between students' performance on the high-stakes test and students' test anxiety classification. By including students classified as low, moderate, and high test-anxious, it is possible to examine performance differences between moderately test-anxious children and low and high test-anxious children.

Test Performance Differences across Test Anxiety Levels

As hypothesized, there was a significant association between student performance on the MEAP and students' test anxiety classification. Figure 5 illustrates students' average English Language Arts and Math MEAP performance based on students' test anxiety classifications. When students' CTAS-Total scale scores were used to classify students as low, moderate, or highly test-anxious, students classified as low test-anxious outperformed students classified as moderately test-anxious on both the English Language Arts and Math sections of the MEAP. Low test-anxious students also outperformed students classified as highly test-anxious on the Math section of the MEAP. Similarly, when students' TASC scale scores were used to classify students as low, moderate, or highly test-anxious, students classified as low test-anxious outperformed students classified as moderately and highly test-anxious on both the English Language



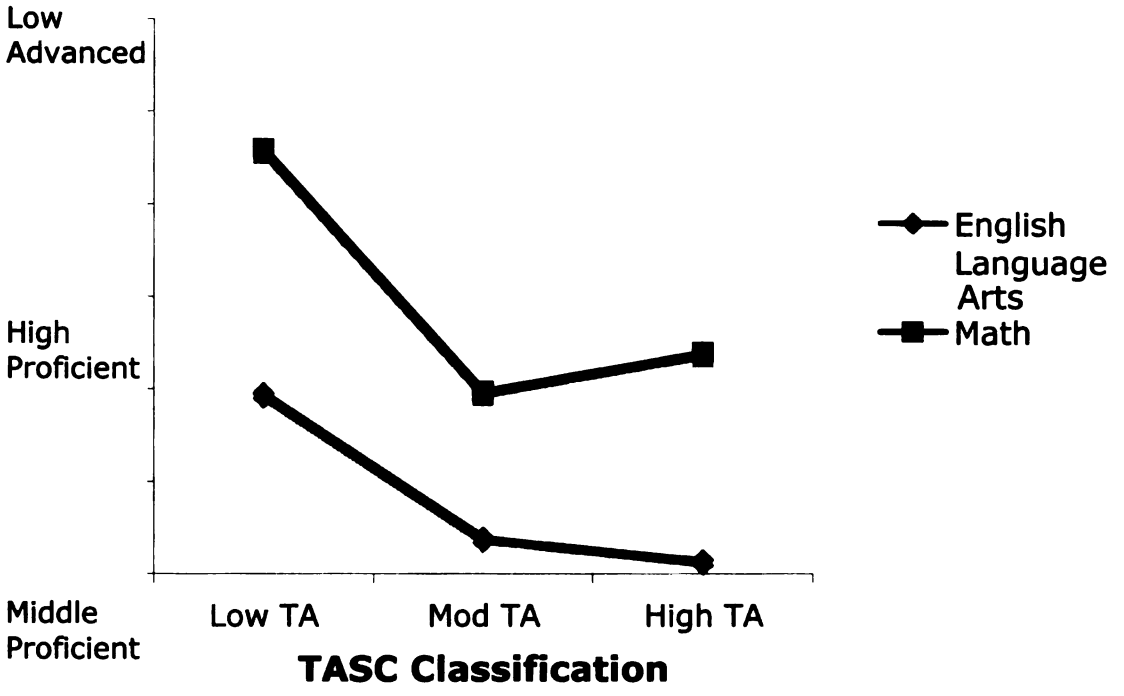


Figure 5: Test Performance Levels by Test Anxiety Classification

Arts and Math sections of the MEAP. There were no differences found between moderately and highly test-anxious students' performance on either the Math and English Language Arts sections of the MEAP using both classification systems.

These findings have practical significance because they show that students' test performance is associated with heightened test anxiety. Perhaps most significant is the finding that students' classified as moderately and highly test-anxious performed significantly less well on the MEAP than students classified as low test-anxious. This finding highlights the importance of including moderately test-anxious students in the analyses as these students experienced performance impairments. Combined with the finding that significantly more students reported moderate test anxiety in relation to high-stakes testing than classroom testing, this study suggests that up to 68% of students experience heightened test anxiety and that may be associated with impaired performance on high-stakes tests in comparison to their low test-anxious peers.

The findings of the current study are supported by previous research examining the relationship between performance and test anxiety. In a meta-analysis of students from third grade to postsecondary school, Hembree (1988) found that reading and mathematics achievement was significantly and inversely related to test anxiety (r = -0.24 and r = -0.22, respectively). Similarly, among a large sample of elementary school students, Hill and Sarason (1966) found that the strength of the negative correlation between test anxiety and achievement increased over the elementary school years. Alternatively, the findings are only partially supported by the only other study in the literature that specifically examines differences in student performance in relation to test anxiety classification. Putwain (2008) found that when students were dichotomized into

low and high test-anxious groups, high test-anxious students performed significantly lower than their low test-anxious peers on low and medium-stakes examinations, while performing slightly better on a high-stakes examination. However, differences in how low and high test-anxious students performed on three examinations in the Putwain study may be related to how students perceived the actual stakes of the examinations, making it difficult to compare these findings directly with the current study findings. Ultimately, the current study suggests that moderate and high test anxiety was associated with significantly impaired test performance on the high-stakes assessment, indicating that reclassification from low to moderate or high test anxiety may have significant educational implications for students. In light of this finding, it is necessary to understand how influential test anxiety was in predicting test performance on the high-stakes assessment.

Association between Test Performance, Test Anxiety, and Student Characteristics

Using multiple regression analyses, the current study examined the unique influence that test anxiety, student grades, and student demographic characteristics had in predicting student test performance on the different content areas of the MEAP. The unique variance in English Language Arts scores explained by student test anxiety ranged from 1.3% to 3.7% across grades three through five. The unique variance in Math scores explained by student test anxiety ranged from 3.2% to 10.7% across grades three through five, suggesting that students' math performance was more strongly associated with test anxiety. Finally, the unique variance in Science scores explained by student test anxiety was 1.2% for students in grade five. When student scholastic achievement was controlled for, the unique contribution of test anxiety in predicting test performance was significant

for third and fifth grade students' Math scores and fourth grade students' English Language Arts scores. Test anxiety was also marginally significant in predicting fourth grade students' math scores and fifth grade students' English Language Arts scores.

Therefore, the hypothesis that test anxiety would be a significant predictor of test performance even when controlling for student scholastic achievement was partially supported. Across all analyses, test anxiety accounted for between 1.2 and 10.7% of the variance in students' test performance, suggesting that although test anxiety was a significant predictor of some MEAP outcome measures across grades, the overall impact that test anxiety had on students' test performance was small. The finding that test anxiety significantly contributed to the model of MEAP test performance for some outcome measures differs from a previous study indicating that pressure and test anxiety did not significantly contribute to test performance (Mulvenon et al., 2005). Alternatively, Mulvenon and colleagues' finding that test anxiety explained only a relatively small amount of the variance in test performance was supported. In sum, although the unique contribution of test anxiety to test performance was small in the current study, test anxiety significantly contributed to students' test performance.

In the current study, test anxiety contributed significantly to grade three students' Math MEAP scores and grade three and four students' English Language Arts MEAP scores, even after student achievement, student sex, ethnicity, SES, and educational verification were controlled for. This finding provides further evidence that the impact of test anxiety on student test performance cannot be explained away by student scholastic achievement or student demographic variables. When examining the contribution of test anxiety in predicting test performance across grade levels, it is of interest that as student

grade increased, the relationship between student test anxiety and student test performance decreased. This finding differs from a longitudinal study that found the strength of the negative correlation between test anxiety and achievement increased over the elementary school years, with achievement becoming significantly more impaired by test anxiety as students advanced in elementary school (Hill & Sarason, 1966). In the current study, it is possible that as students become more experienced with high-stakes testing in the later elementary school years, the interference experienced as a result of test anxiety decreases. Further studies examining the developmental nature of the relationship between test anxiety, performance, and student grade level are needed.

Association between Test Anxiety and Student Characteristics

This study also examined the relationship between test anxiety and student demographic characteristics. Previous studies of test anxiety suggest that there is a significant relationship between test anxiety and child characteristics, including age, sex, race, and socioeconomic status (Hembee, 1988). The current study found that there were significant differences in test anxiety in relation to the high stakes assessment when students were classified on the basis of their sex and grade. Student test anxiety on classroom tests was also significantly different when students were classified by sex but not by grade.

Contrary to the hypothesis that student characteristics, including ethnicity, educational verification, and socioeconomic status would significantly predict test anxiety on the high-stakes assessment since testing scores are reported using these categories, the current study found that these variables did not significantly predict test anxiety in relation to the high-stakes testing situation. The finding that there was no

relationship between student test anxiety and student ethnicity differs from previous studies that suggest ethnic minority students report significantly greater test anxiety than Caucasian students on high-stakes tests (Putwain, 2007) and testing in general (Turner et al., 1993; Wren & Benson, 2004). Additionally, a meta-analysis examining the effect of ethnicity and age on test anxiety suggests that African-American children report significantly more test anxiety than Caucasian children, with greater differences reported in the elementary years, grades 2 - 4: Effect size (ES) = 0.52 and grades 5 - 8: ES = 0.21 (Hembree, 1988). Similarly, the finding that there was no relationship between student test anxiety and student SES differs from the Putwain (2007) finding found that teenagers from low socioeconomic backgrounds reported significantly more test anxiety on high-stakes tests than students from high socioeconomic backgrounds.

The current study finding of sex differences in reported test anxiety is substantiated by numerous research studies examining differences in females and males' reports of test anxiety in relation to testing in general (Hembree, 1988) and high-stakes testing (Putwain, 2007). A meta-analysis of the literature suggests that gender differences in test anxiety are also influenced by child age, with moderate differences in grades 5-10 (ES = 0.43) and small differences in lower and upper grades, (grades 1 and 2: ES = 0.14; grades 3 and 4: ES = 0.28; and grades 11, 12, and postsecondary: ES = 0.27; Hembree, 1988). In the current study, the effect size of the difference in test anxiety reported by males and females was similar across both high-stakes and classroom testing situations and all measures, ranging from *r*s = -.21 to -.27. These effect sizes closely replicate the findings of the 26 studies of students in grades three and four that were examined in the meta-analysis conducted by Hembree.

Interestingly, in the current study, student grade was associated with differences in test anxiety in relation to high-stakes testing but not to classroom testing situations. The current study findings are partially supported by previous research findings on the effect of student grade on test anxiety, which have examined students' responses to testing situations in general rather than to high-stakes testing situations in particular. Studies of the effect of grade on test anxiety suggest that test anxiety increases over the course of elementary school (Hembree, 1988, Hill & Sarason, 1966; McDonald, 2001; Sarason et al., 1960). According to a meta-analysis examining differences in test anxiety between students in adjacent grades, test anxiety increased through fifth grade before remaining relatively constant through the end of high school (Hembree). Hembree found that there while there were increases in anxiety between adjacent grades 3 to 4 and 4 to 5, the increase was only significant in grade 3 to 4 (ES = -.24, $p < .01$). Thus, the current study finding that grade was significantly related to test anxiety on the high-stakes assessment, with students in grade four reporting greater test anxiety than students in grade three on both the CTAS-Total and BASC-2-TA scales, is supported by the meta-analytic findings. Alternatively, the finding that grade four students were also more anxious than students in grade five on the CTAS-Total scale is not supported by prior research, or the findings of no difference in test anxiety as measured by the BASC-2-TA scale between students in these grades.

The finding in the current study that student grade was not associated with test anxiety in relation to classroom testing differs from the meta-analysis findings. This difference is of interest in light of the significant differences found between grade and test anxiety on the high stakes assessment. Previous researchers have suggested that

increases in test anxiety over elementary school may be related to increased exposure to testing over time and to conditioning that occurs following repeated concern about test failure, pressure to perform well, or feedback about poor performance (King & Ollendick, 1989). While students in the current study had frequent experience classroom testing, including weekly spelling tests, tests on math facts, and tests on curriculum content, they were participating in high-stakes testing for the first, second, or third time, depending upon their grade level. Thus, the increase in test anxiety reported from grade three to four may have resulted from the MEAP having more salience for fourth grade students following the experience of taking the assessment for the first time in third grade. Further research is needed to clarify the relationship between student grade and test anxiety between typical and high-stakes testing situations as well as the developmental trajectory of test anxiety over the course of children's schooling.

Finally, regression analyses that examined the unique contribution of demographic variables in predicting student test anxiety suggest that although student sex, grade, and SES contributed significantly to students' test anxiety scores on some measures of classroom tests and high-stakes tests, the total variance in test anxiety explained by any one of these demographic variables was quite small, ranging from 1 to 7%. This finding suggests that student demographic variables alone are not good predictors of which students will report high levels of test anxiety in relation to classroom testing or high-stakes testing. Thus, student characteristics cannot be used alone to identify students at risk for experiencing significant test anxiety, and direct evaluation of student test anxiety is warranted.

Measures of Test Anxiety

Correlational analyses between the three different measures used to assess test anxiety in relation to high-stakes and classroom testing conditions indicated that there was high convergence between all measures, all $r_s > .70$. The high convergence between all three measures suggests that all are acceptable measures of the psychological construct of test anxiety. As the gold standard measure of test anxiety, the strong relationships identified between the TASC scale and the CTAS-Total scale and the BASC-2-TA subscale supports the concurrent validity of the CTAS-Total scale and the BASC-2-TA subscale as measures of test anxiety. It is important to note, however, that the concurrent validity of these scales as measures of test anxiety does not provide any indication about the validity of the classification systems used to sort students into low, moderate, or high test-anxious groups based on TASC and CTAS-Total scores.

The biopsychosocial model of test anxiety hypothesizes that test anxiety includes components of cognitive interference, psychological arousal, and behavioral interference (Lowe et al., 2008; Zeidner, 1997). The CTAS scale is the only publicly available measure designed to assess test anxiety in young children that specifically examined all of these components of test anxiety and provides subscale scores in each of these areas. Examination of the correlations between these subscales and the BASC-2-TA subscale and the TASC scale provide an indication of how well these measures assess these three components of test anxiety. There were strong correlations between the CTAS subscales assessing cognitive and physiological aspects of test anxiety and the BASC-2-TA as well as the TASC scale, $.64 \leq r_s \leq .74$. However, the relationship between the CTAS subscale measuring behavioral interference and the BASC-2-TA as well as the TASC scale was

weaker, $.41 \leq r_s \leq .50$. These results suggest that the BASC-2-TA and TASC scales are better at assessing cognitive and physiological aspects of test anxiety than behavioral aspects of test anxiety. Thus, these measures are more reflective of the two component, worry and emotionality conceptualization of test anxiety advanced by Liebert and Morris in 1967.

Additional consideration of the correlation between the CTAS-Total and the BASC-2-TA scales across each testing condition is warranted. One potentially important use of a test anxiety scale could be screening student test anxiety in relation to both typical classroom testing and high-stakes testing to identify students who might benefit from interventions designed to decrease test anxiety. While two assessments could be conducted in an applied educational setting, the need for repeated testing could be eliminated by identifying a test anxiety scale that assesses typical classroom testing and is also highly correlated with test anxiety associated with high-stakes testing. In the current study, while both the CTAS-Total and the BASC-2-TA measures were significantly correlated across the testing conditions, the strongest relationship between testing conditions was on the CTAS-Total scale. Therefore, the best predictor of students' test anxiety in relation to the high-stakes assessment was students' CTAS-Total scale score in relation to classroom testing.

Thus, in light of current theoretical conceptualizations of test anxiety and the findings of the current study, the CTAS appears to be the most appropriate measure of the three used in the current study for assessing biopsychosocial aspects of test anxiety in young children in relation to both classroom testing in general and high-stakes testing. While the TASC has been the most extensively used measure of test anxiety in the

literature, the current study suggests that the TASC is only moderately correlated with a measure of behavioral interference. Coupled with concerns about the length and reading level of TASC items (Wren & Benson, 2004), use of the CTAS scale to assess test anxiety appears most appropriate. Alternatively, when the goal of assessment is screening overall test anxiety efficiently and universally within an educational setting, the brevity of the 7-item BASC-2-TA, coupled with the strong correlation found between the BASC-2-TA and the CTAS-Total scale make the BASC-2-TA an attractive and appropriate option.

Limitations

There were several limitations related to the student sample that confine the generalizability of this study's findings. First, this study design involved the examination of how students from three schools in one Michigan school district responded to classroom testing and high-stakes testing conditions. As such, the results of this study may not generalize to schools in other districts in Michigan. Within states, districts and schools vary in how they approach test preparation and administration, which may affect students' responses to high-stakes testing. Additionally, since high-stakes assessments designed to meet NCLB requirements vary from state to state in multiple ways, including, but not limited to the mode of administration, the length, and the consequences attached to the assessments, caution must be exercised in using the current study findings to draw conclusions about the how students respond to high-stakes testing in other states. Similarly, the overall sample size of this study is extremely small in comparison to the number of children enrolled in public schools in the United States who take high-stakes assessments to comply with NCLB requirements.

Although the power to detect small to moderate effect size differences between student sex, grade, ethnicity, SES, and educational verification, test anxiety, and test performance was sufficient, further research with a larger, more diverse population of students is needed to clarify the relationship between these demographic variables, test anxiety, and test performance in high and low stakes testing situations in young children. Similarly, further research with a larger sample of students at each grade level is needed to clarify the relationship between test anxiety, scholastic achievement, demographic characteristics, and test performance on high-stakes assessments. Additionally, a limitation of this study was the reliance on parent-report to measure student socioeconomic status and educational verification. This aspect of the study design may have resulted in somewhat less accurate reports of student demographic characteristics. Using school data to identify students' Free and Reduced Lunch and educational verification status in future studies may result in more accurate demographic information about students.

Finally, there were methodological limitations with the teacher perception data and caution is warranted in interpreting these findings. In addition to lower teacher participation rates during the second survey, the teacher data were not coded in a way that allowed for direct comparisons between teachers' perceptions of classroom testing and high-stakes testing. Therefore, it was necessary to analyze the data using independent t-tests rather than repeated measures t-tests. Future studies of perception data should include paired data. Additionally, by pairing teacher perception data with student data, it would be possible to examine whether or not individual classroom teachers' anxious responses to testing or perceptions of students' test anxiety affect the performance of

their students or students' self-reported test anxiety. In sum, this study provided a number of unique additions to the literature on test anxiety among young children in relation to high-stakes testing and classroom testing; however, caution must be used in interpreting this study's findings given the aforementioned methodological limitations.

Clinical and Educational Implications

Several clinical and educational implications related to identifying children at risk for impairing levels of test anxiety can be gleaned from this study. Perhaps the most important finding in this study was the significant difference in test anxiety reported by students between high-stakes and classroom testing. The results of this study suggest that as a whole, students report significantly more overall test anxiety and more cognitive and physiological symptoms of test anxiety in response to high-stakes testing than classroom testing. This increase was associated with more students reporting moderate levels of test anxiety in the high-stakes condition than the classroom testing condition. This study was the first to specifically examine differences in students' test anxiety responses to different testing conditions, and it adds substantially to the test anxiety literature by demonstrating that students perceive these two types of testing differently. This difference has implications for teachers and educators involved in student preparation for high-stakes tests. By understanding that students may experience different test anxiety symptoms across different testing conditions, teachers and educators can more effectively prepare students for these different types of tests.

Additionally, the current study found that students who reported moderate and high test anxiety in relation to high-stakes testing had significantly lower test performance in comparison to low test-anxious students. This finding suggests that

performance impairments are associated with moderate *and* high test anxiety and that there are no differences in performance between moderately and highly test-anxious students. Hence, the current study provides evidence supporting concerns reported by teachers that high stakes testing programs result in increased stress, worry, and negative outcomes for students (Abrams et al., 2003; Barksdale-Ladd & Thomas, 2000; Jones & Egley, 2004; 2006; Jones et al., 1999).

There is a need for teachers and educators to consider test anxiety during test preparations given that students report more test anxiety in response to high-stakes testing and that heightened test anxiety is associated with impaired testing performance. Educators need to be knowledgeable about intervention programs shown to effectively decrease students' test anxiety (e.g., Beidel, Turner, & Taylor-Ferreira, 1999; Ergene, 2003; Tryson, 1980). The finding that students report more test anxiety in relation to high-stakes testing provides a rationale for universally teaching students some coping strategies to reduce test anxiety symptoms prior to taking high-stakes tests. Interventions to reduce test anxiety may be appropriate for the 60 - 70% of the elementary school student population who are classified as moderately to highly test anxious in relation to high-stakes testing.

The findings of the current study also have implications for educators interested in developing efficient systems for identifying which students would most benefit from more intensive interventions to reduce test anxiety. While the CTAS is a more theoretically sound instrument that assesses biopsychosocial aspects of student test anxiety, the BASC-2-TA subscale is highly correlated with this measure and it is significantly more brief (7 versus 30 items). The brevity of the BASC-2-TA subscale

may make universal test anxiety screening more feasible and acceptable in schools. Additionally, given the movement towards conducting universal mental health assessments and interventions in schools (Salmon & Kirby, 2008), the use of the complete BASC scale could provide educators with a measure of test anxiety as well as indications of student functioning across internalizing and externalizing areas as well.

As experts in data-based decision-making, school psychologists are uniquely prepared to organize, execute, and interpret universal screening programs in the school setting. The response to intervention movement has increased school psychologists' involvement in universal screenings of students' academic skills and targeted progress monitoring of students' who are identified as having academic skill deficits. These skills have been shown to transfer to effective universal social-emotional screening and intervention programs (e.g., Cheney, Flower, & Templeton, 2008). Similarly, universal screenings of test anxiety and progress monitoring the effectiveness of any programs designed to decrease students' test anxiety symptoms in response to different testing conditions could be completed by school psychologists.

Future Research

The results of this study suggest that as a whole, elementary school children experience significantly more test anxiety in relation to high-stakes assessments than classroom testing. Examination of individual students' test anxiety classifications across testing conditions revealed that 60% of students reported stable test anxiety levels across testing conditions, 25% reported significant increases in test anxiety level, and 15% of students reported significant decreases in test anxiety. These findings add nuance to previous research, which have been limited by methodological weaknesses and report

equivocal findings on children's responses to high-stakes testing. Thus, this study adds to this limited research base and suggests that while as a whole, students report significantly more test anxiety in relation to high-stakes assessments, individual students had differing responses to the different testing conditions. Further research is needed to clarify which students are most at risk for experiencing increased test anxiety in relation to high-stakes testing and how these students differ from those who reported less test anxiety in relation to high-stakes testing.

The finding that a large proportion of students (42%) who reported low test anxiety in relation to classroom testing also reported moderate test anxiety in relation to high-stakes testing takes on particular significance in light of the finding that moderately test-anxious students performed significantly less well on the high-stakes assessment than low test-anxious students. Future research studies that expand upon the current study findings and specifically examine the effect of low, moderate, and high test anxiety on test performance are needed. It will also be important for researchers to expand upon the findings of the current study and existing literature to clarify the relationship between increased test anxiety, test performance, and testing stakes. Further research that specifically examines differences between students' test anxiety responses to testing with varying stakes is needed. Moreover, future research should examine whether or not the impact that test anxiety has on testing performance differs depending upon the stakes of the test. Similarly, replications and extensions of the current study are needed to clarify whether or not increases in test anxiety on annual tests required by NCLB tests exist across different states and more diverse samples of students. Similarly, the relationship between test anxiety and test performance needs to be examined across more diverse and

larger samples of students. Future studies should involve students from different states, districts with and without test preparation programs, and diverse socioeconomic and ethnic backgrounds.

Additionally, the biospychosocial model of test anxiety supports the construct validity of the Children's Test Anxiety Scale (CTAS). However, as a recently developed measure that has only been studied by the test developer, additional examination by an independent researcher is needed to verify the psychometric properties of the CTAS. Similarly, there is a need to develop a classification system for the CTAS that can be used to identify children at risk for performance impairments or significant psychological distress in relation to testing. The current suggested cutpoints for low, moderate, and high test anxiety are based on the assumption that test anxiety is a normally distributed construct; however, findings from the current study suggest that test anxiety is a positively skewed construct. Therefore, identification of cutpoints based on verification of impairment using a measure of test performance would be more appropriate than using scores above or below one standard deviation of the mean of the normative sample to determine test anxiety classification. Similarly, further examination of the feasibility, acceptability, and utility of the CTAS and BASC-2-TA as measures of student test anxiety are needed.

Finally, test anxiety intervention research is needed to develop effective interventions for test-anxious students that can be practically implemented into educational settings. Future research examining the effectiveness, feasibility, and acceptability of interventions designed to decrease students' test anxiety and increase students' coping skills are needed. Studies are needed that not only examine how

effective test anxiety interventions are in reducing test anxiety, but also examine how effective test anxiety interventions are in decreasing performance impairments associated with test anxiety.

APPENDICES

Appendix A:

Demographic Information Form

DEMOGRAPHIC INFORMATION FORM

Child Name: _____ Child's Date of birth: _____

Sex of Child: ___ Male Child's Age: _____

___ Female Child's Grade: _____

Child's ethnicity/race:

___ Caucasian ___ African American

___ Mexican, Mexican-American ___ Other Latino or Hispanic

___ American Indian ___ Asian

___ Pacific Islander ___ Other (describe)_____

Does your child receive **special educational services** for have any of the following? (Please circle)

Learning problem	Yes / No	Vision or hearing problem	Yes / No
Cognitive Delay	Yes / No	Language Delay	Yes / No
ADHD	Yes / No	Physical handicap	Yes / No
Emotional/behavioral problem	Yes / No		

Caregiver Information (i.e. mother, father, stepparent, adoptive parent, partner, etc. who are responsible for caring for the child):

Caregiver #1

Highest level of education completed:

___ Grades 0-8

___ Grades 9-11

___ High School or GED

___ Some college (1 year or more)

___ College graduate

___ Post-college

Current Employment:

___ Yes, full time (describe)_____

___ Yes, part time(describe)_____

___ Not working (receiving gov. assistance)

___ Not working by choice

Caregiver #2

Highest level of education completed:

___ Grades 0-8

___ Grades 9-11

___ High School or GED

___ Some college (1 year or more)

___ College graduate

___ Post-college

Current Employment:

___ Yes, full time (describe)_____

___ Yes, part time(describe)_____

___ Not working (receiving gov.assistance)

___ Not working by choice

Appendix B:

Academic Performance Form

ACADEMIC PERFORMANCE FORM

Third-Grade Grades from Fall 2008

Reading:	Level	_____
	Decoding	_____
	Comprehension	_____
	Fluency	_____
Writing:	Ideas	_____
	Organization	_____
	Word Choice	_____
	Conventions	_____
	Spelling	_____
Mathematics:	Numbers/Operations	_____
	Measurement	_____
	Geometry	_____
	Data/Problem-solving	_____

Fourth and Fifth-Grade Grades from Fall 2008

Reading	_____
Writing	_____
Mathematics	_____
Science (5th only)	_____

MEAP Scores and Levels

Reading	_____
Writing	_____
ELA	_____
Mathematics	_____
Science (5th only)	_____

DIBELS

Benchmark 1 ORF _____

Appendix C:

Teacher High-Stakes Testing Questionnaire

On this page there are a series of questions about your perceptions of MEAP testing. Your responses to these questions are an indication of your consent for this information to be used as part of a research study examining test anxiety among elementary school children. You have the right to not participate in this study or to not answer any individual questions. All information will be kept confidential and no identifying information will be linked to this data.

School Name _____

Grade Taught _____

In general,	**NOT AT ALL**	**SOMEWHAT**	**QUITE A BIT**	**VERY**
How *anxious* were your students about the MEAP assessment before they took it?	1	2	3	4
How *anxious* were your students about the MEAP assessment during the assessment?	1	2	3	4
How *anxious* were you about how well your students would do on the MEAP assessments?	1	2	3	4

In general,	**Very much below their ability**	**Somewhat below their ability**	**At their ability**	**Somewhat above their ability**	**Very much above their ability**
How well do you think your students performed on the MEAP assessment?	1	2	3	4	5

Comments:

Appendix D:

Teacher Classroom Tests Questionnaire

On this page there are a series of questions about your perceptions of classroom testing practices. Your responses to these questions are an indication of your consent for this information to be used as part of a research study examining test anxiety among elementary school children. You have the right to not participate in this study or to not answer any individual questions. All information will be kept confidential and no identifying information will be linked to this data.

School Name _____

Grade Taught _____

For the questions that follow, class tests are defined as any graded assessment of scholastic skills that occurred during the past 3 weeks. This may include spelling tests, quizzes or tests in content areas (such as math, science, social studies), or any other formal assessment of student learning.

1. Please list the names of all class tests that have occurred in the last 3 weeks

_____

_____

_____

In general,	NOT AT ALL	SOMEWHAT	QUITE A BIT	VERY
2. How *anxious* were your students about class tests before they took them?	1	2	3	4
3. How *anxious* were your students about class tests during the tests?	1	2	3	4
4. How *anxious* were you about how well your students would do on classroom tests?	1	2	3	4

In general,	Very much below their ability	Somewhat below their ability	At their ability	Somewhat above their ability	Very much above their ability
5. How well do you think your students performed on classroom tests?	1	2	3	4	5

Comments:

Appendix E:

Parental Information Letter

MICHIGAN STATE
UNIVERSITY

September ___, 2008

Dear Parent/Guardian:

We are writing to invite your child to participate in a research study examining how students feel about test taking. We have partnered with _____ Elementary School and all of the third, fourth, and fifth grade students at _____ are being invited to participate in this study.

We hope to learn how elementary school students feel about test taking and in particular, how they feel about the MEAP. We also hope to learn if students' feelings impact their performance on tests.

The study will involve your child completing two brief surveys about his/her feelings about tests. The surveys will be completed during non-academic time and your child will receive a small prize and snack as a thank you. The surveys will be completed for research purposes only and the results will be kept confidential.

We have enclosed a consent form for you to fill out in order for your child to participate in this research study. Please complete the consent and demographic forms and return them to your child's classroom teacher by _____, 2008.

If you have any questions, please feel welcome to contact the researchers, Natasha K. Segool, by phone: 517-231-9116 or email: segoolna@msu.edu, or John S. Carlson, by phone: 517-432-0843; email: carlsoj@msu.edu.

Thank you,

Natasha K. Segool, M.A. _____

John S. Carlson, Ph.D. _____

COLLEGE OF
EDUCATION
Department of Counseling, Educational Psychology, and Special Education

Natasha K. Segool, M.A.
Michigan State University
401C Erickson Hall
East Lansing, Michigan
48824-1034

segoolna@msu.edu

Appendix F:

Consent Form

Research Participant Information and Consent Form

Your child is being asked to participate in a research project. Researchers are required to provide a consent form to inform you about the study, to convey that participation is voluntary, to explain risks and benefits of participation, and to empower you to make an informed decision. You should feel free to ask you the researchers any questions may have.

Study Title: Test Anxiety Associated with High-Stakes Testing among Elementary School Children: Prevalence, Predictors, and Relationship to Student Performance

Researcher and Title: John Carlson, Ph.D., Associate Professor
Natasha Segool, M.A, Doctoral Candidate

Department and Institution: Department of Counseling, Educational Psychology, and Special Education, Michigan State University

PURPOSE OF RESEARCH:
Your child is being asked to participate in a research study examining how students feel about test taking and whether or not these feelings affect their performance on tests. Students are being asked how they feel about taking classroom tests and the Michigan Educational Assessment Program (MEAP) tests. Your child has been selected as a possible participant in this study because he/she is taking the MEAP test this fall. All of the third, fourth, and fifth grade students at your child's school are being invited to participate in this study. In the entire study, approximately 500 students are being asked to participate in the study. Your child's participation in this study will take about a total of forty minutes. From this study, the researchers hope to learn how elementary school students feel about test taking and in particular, how they feel about the MEAP. We also hope to learn if students' worries impact their performance on tests.

WHAT YOU AND YOUR CHILD WILL DO:
You will be asked to fill out a brief demographic form about your child. Your child will complete two surveys about his/her feelings about tests. The first survey will be completed after your child finishes taking the MEAP. The second survey will be completed about three weeks later. Each survey will take fifteen to twenty minutes to complete. The surveys will be read aloud to students in the classroom and students will complete the surveys individually. The MEAP is a routine part of your child's academic work. The researchers will collect data on your child's performance on this assessment as well as your child's academic performance in the prior academic year. The surveys will be completed for research purposes only and the results will not be shared with you, your child, the teacher, or the school.

POTENTIAL BENEFITS:
Your child will not directly benefit from your participation in this study, with the possible exception that answering the questions could be a positive experience for you child. Your child's participation in this study may contribute to the understanding of how children feel about testing. This research, along with future research, may increase our knowledge of test anxiety and its relationship to school achievement, thus potentially benefiting children with test anxiety in the future. Following the completion of the study, you will receive information on the result of the study if you so wish. These will be aggregated results, and not results of your child as an individual.

POTENTIAL RISKS:
This study poses no more than minimal risk for your child, although there is the potential for psychological discomfort. We will ask your child to complete surveys about his/her thoughts, feelings, and behaviors before, during, and after tests. Answering some of the questions may cause your child to experience discomfort or distress. The researchers will be available to answer children's questions during and after the surveys.

PRIVACY AND CONFIDENTIALITY:
The data for this project will be kept confidential to the greatest extent allowable by law. It is important for you to know, however, that the researchers are required to report evidence of child abuse. Although the researchers will know your child's identify and your child's name will be on some documentation during the study, all identifying information will be removed as soon as the study is complete. At that time, an identification number will be assigned to your data and your child's name will be removed from all paperwork. Completed surveys will be kept in a locked file cabinet in the office of one of the researchers. All documents will be destroyed ten years after study completion. The results of this study may be published or presented at professional meetings, but the identities of all research participants will remain anonymous. It will not be possible for readers to know who participated in the study.

YOUR RIGHTS TO PARTICIPATE, SAY NO, OR WITHDRAW
Participation in this research project is completely voluntary. You and your child have the right to say no. You or your child may change your minds at any time and withdraw from the study. You or your child may also choose not to answer specific questions or to stop participating at any time. Whether or not you or child chooses to participate will have no effect on your child's grade.

COSTS AND COMPENSATION FOR BEING IN THE STUDY:
It does not cost anything to participate in this study. Your child will receive a small prize/snack following the completion of each survey.

CONTACT INFORMATION FOR QUESTIONS AND CONCERNS
If you have any questions about this study, such as scientific issues, how to do any part of it, or to report an injury, please contact the researchers, Natasha Segool, by phone: 517-231-9116; email: segoolna@msu.edu or John S. Carlson, by phone: 517-432-0843; email: carlsoj@msu.edu or regular mail: 431 Erickson Hall, East Lansing, MI 48824. If you have questions or concerns about your role and rights as a research participant, or would like to register a complaint about this study, you may contact, anonymously if you wish, the Michigan State University's Human Research Protection Program at 517-355-2180, Fax 517-432-4503, or e-mail irb@msu.edu or regular mail at 202 Olds Hall, MSU, East Lansing, MI 48824

DOCUMENTATION OF INFORMED CONSENT.
Please select a box, fill in your child's name, and sign below.

☐ Yes, my child _____ may participate in this research study

☐ No, my child _____ may not participate in this research study.

_____ _____
Parent Signature Date

Appendix G:

Survey One

Survey One Code # _____

Student Name _____

School _____

Teacher Name _____

Grade _____

We are from Michigan State University. We would like to learn more about children's feelings about tests. To do this, we are asking you and other children to take part in a research study. In this study, we will be asking you some questions about your feelings and behaviors.

If you agree to be part of the study, you will be asked to answer some questions that will take about 20 minutes. You do not have to answer any of the questions if you do not want to or if they make you feel uncomfortable.

No one will know the information belongs to you and the results of the research will not be able to be connected to you. Only information that you tell us that we think is dangerous to your health or others will be shared with your parents or others who can help keep you or others safe.

Name _____ Date _____

Survey One Code # _____

How I Feel about the MEAP

Directions: Listen to each question as it is read aloud. Circle one answer for each question.

While I am taking the MEAP . . .	**ALMOST NEVER**	**SOME OF THE TIME**	**MOST OF THE TIME**	**ALMOST ALWAYS**
1. I wonder if I will pass.	1	2	3	4
2. My heart beats fast.	1	2	3	4
3. I look around the room.	1	2	3	4
4. I feel nervous.	1	2	3	4
5. I think I am going to get a bad score.	1	2	3	4
While I am taking the MEAP . . .	**ALMOST NEVER**	**SOME OF THE TIME**	**MOST OF THE TIME**	**ALMOST ALWAYS**
6. It is hard for me to remember the answers.	1	2	3	4
7. I play with my pencil.	1	2	3	4
8. My face feels hot.	1	2	3	4
9. I worry about failing.	1	2	3	4
10. My belly feels funny. . . .	1	2	3	4
While I am taking the MEAP . . .	**ALMOST NEVER**	**SOME OF THE TIME**	**MOST OF THE TIME**	**ALMOST ALWAYS**
11. I worry about doing something wrong.	1	2	3	4
12. I check the time.	1	2	3	4
13. I think about what my score will be. .	1	2	3	4
14. I find it hard to sit still.	1	2	3	4
15. I wonder if my answers are right.	1	2	3	4

Survey One Code # _____

While I am taking the MEAP . . .	**ALMOST NEVER**	**SOME OF THE TIME**	**MOST OF THE TIME**	**ALMOST ALWAYS**
16. I think that I should have studied more.	1	2	3	4
17. My head hurts.	1	2	3	4
18. I look at other people. . . .	1	2	3	4
19. I think most of my answers are wrong.	1	2	3	4
20. I feel warm.	1	2	3	4
While I am taking the MEAP . . .	**ALMOST NEVER**	**SOME OF THE TIME**	**MOST OF THE TIME**	**ALMOST ALWAYS**
21. I worry about how hard the MEAP is.	1	2	3	4
22. I try to finish up fast.	1	2	3	4
23. My hand shakes.	1	2	3	4
24. I think about what will happen if I fail.	1	2	3	4
25. I have to go to the bathroom. .	1	2	3	4
While I am taking the MEAP . . .	**ALMOST NEVER**	**SOME OF THE TIME**	**MOST OF THE TIME**	**ALMOST ALWAYS**
26. I tap my feet.	1	2	3	4
27. I think about how poorly I am doing. .	1	2	3	4
28. I feel scared.	1	2	3	4
29. I worry about what my parents will say. .	1	2	3	4
30. I stare.	1	2	3	4

	TRUE	FALSE
31. I hate taking the MEAP. . .	T	F
32. I worry about the MEAP more than my classmates do. . . .	T	F
33. No matter how much I study for the MEAP, I am afraid I will fail. .	T	F

	NEVER	SOMETIMES	OFTEN	ALMOST ALWAYS
34. Even when I try hard, I fail.	N	S	O	A
35. I get nervous when I take the MEAP.	N	S	O	A
36. When I take the MEAP, I can't think.	N	S	O	A
37. I have trouble sleeping the night before the MEAP.	N	S	O	A

Appendix H:

Survey Two

Survey Two Code # _____

Student Name _____

School _____

Teacher Name _____

Grade _____

We are from Michigan State University. We would like to learn more about children's feelings about tests. To do this, we are asking you and other children to take part in a research study. In this study, we will be asking you some questions about your feelings and behaviors.

If you agree to be part of the study, you will be asked to answer some questions that will take about 20 minutes. You do not have to answer any of the questions if you do not want to or if they make you feel uncomfortable.

No one will know the information belongs to you and the results of the research will not be able to be connected to you. Only information that you tell us that we think is dangerous to your health or others will be shared with your parents or others who can help keep you or others safe.

Name _____ Date _____

Survey Two Code # _____

How I Feel about Tests

Directions: Listen to each question as it is read aloud. Circle one answer for each question.

While I am taking tests . . .	**ALMOST NEVER**	**SOME OF THE TIME**	**MOST OF THE TIME**	**ALMOST ALWAYS**
1. I wonder if I will pass.	1	2	3	4
2. My heart beats fast.	1	2	3	4
3. I look around the room.	1	2	3	4
4. I feel nervous.	1	2	3	4
5. I think I am going to get a bad grade.	1	2	3	4
While I am taking tests . . .	**ALMOST NEVER**	**SOME OF THE TIME**	**MOST OF THE TIME**	**ALMOST ALWAYS**
6. It is hard for me to remember the answers.	1	2	3	4
7. I play with my pencil.	1	2	3	4
8. My face feels hot.	1	2	3	4
9. I worry about failing.	1	2	3	4
10. My belly feels funny. . . .	1	2	3	4
While I am taking tests . . .	**ALMOST NEVER**	**SOME OF THE TIME**	**MOST OF THE TIME**	**ALMOST ALWAYS**
11. I worry about doing something wrong.	1	2	3	4
12. I check the time.	1	2	3	4
13. I think about what my grade will be. .	1	2	3	4
14. I find it hard to sit still. . . .	1	2	3	4
15. I wonder if my answers are right. .	1	2	3	4

Survey Two Code # _____

While I am taking tests . . .	**ALMOST NEVER**	**SOME OF THE TIME**	**MOST OF THE TIME**	**ALMOST ALWAYS**
16. I think that I should have studied more.	1	2	3	4
17. My head hurts.	1	2	3	4
18. I look at other people.	1	2	3	4
19. I think most of my answers are wrong.	1	2	3	4
20. I feel warm.	1	2	3	4
While I am taking tests . . .	**ALMOST NEVER**	**SOME OF THE TIME**	**MOST OF THE TIME**	**ALMOST ALWAYS**
21. I worry about how hard the test is. .	1	2	3	4
22. I try to finish up fast.	1	2	3	4
23. My hand shakes.	1	2	3	4
24. I think about what will happen if I fail.	1	2	3	4
25. I have to go to the bathroom. . .	1	2	3	4
While I am taking tests . . .	**ALMOST NEVER**	**SOME OF THE TIME**	**MOST OF THE TIME**	**ALMOST ALWAYS**
26. I tap my feet.	1	2	3	4
27. I think about how poorly I am doing. .	1	2	3	4
28. I feel scared.	1	2	3	4
29. I worry about what my parents will say.	1	2	3	4
30. I stare.	1	2	3	4

	TRUE	FALSE
31. I hate taking tests.	T	F
32. I worry about tests more than my classmates do.	T	F
33. No matter how much I study for a test, I am afraid I will fail.	T	F

	NEVER	SOMETIMES	OFTEN	ALMOST ALWAYS
34. Even when I try hard, I fail.	N	S	O	A
35. I get nervous when I take tests.	N	S	O	A
36. When I take tests, I can't think.	N	S	O	A
37. I have trouble sleeping the night before a big test.	N	S	O	A

38. Do you worry when the teacher says that she is going to ask you questions to find out how much you know?	Yes	No
39. Do you worry about being promoted, that is, passing from the ___ grade to the ___ grade at the end of the year?	Yes	No
40. When the teacher asks you to get in front of the class and read aloud, are you afraid that you are going to make some bad mistakes?	Yes	No
41. When the teacher says that she is going to call upon some boys and girls in the class to do arithmetic problems, do you hope that she will call upon someone else and not on you?	Yes	No
42. Do you sometimes dream at night that you are in school and cannot answer the teacher's questions?	Yes	No
43. When the teacher says that she is going to find out how much you have learned, does your heart begin to beat faster?	Yes	No
44. When the teacher is teaching you about arithmetic, do you feel that other children in the class understand her better than you?	Yes	No
45. When you are in bed at night, do you sometimes worry about how you are going to do in class the next day?	Yes	No
46. When the teacher asks you to write on the blackboard in front of the class, does the hand you write with sometimes shake a little?	Yes	No
47. When the teacher is teaching you about reading, do you feel that other children in the class understand her better than you?	Yes	No
48. Do you worry more about school than other children?	Yes	No
49. When you are at home and you are thinking about your arithmetic lesson for the next day, do you become afraid that you will get the answers wrong when the teacher calls on you?	Yes	No
50. If you are sick and miss school, do you worry that you will do more poorly in your schoolwork than other children when you return to school?	Yes	No
51. Do you sometimes dream at night that other boys and girls in your class can do things you cannot?	Yes	No

52. When you are home and you are thinking about your reading lesson for the next day, do you worry that you will do poorly on the lesson?	Yes No
53. When the teacher says that she is going to find out how much you have learned, do you get a funny feeling in your stomach?	Yes No
54. If you did very poorly when the teacher called on you, would you probably feel like crying even though you would try not to cry?	Yes No
55. Do you sometimes dream at night that the teacher is angry because you do not know your lessons?	Yes No
56. Are you afraid of school tests?	Yes No
57. Do you worry a lot *before* you take a test?	Yes No
58. Do you worry a lot *while* you are taking a test?	Yes No
59. *After* you have taken a test do you worry about how well you did on the test.	Yes No
60. Do you sometimes dream at night that you did poorly on a test you had in school that day?	Yes No
61. When you are taking a test, does the hand you write with shake a little?	Yes No
62. When the teacher says that she is going to give the class a test, do you become afraid that you will do poorly?	Yes No
63. When you are taking a hard test, do you forget some things you knew very well before you started taking the test?	Yes No
64. Do you with a lot of times that you didn't worry so much about tests?	Yes No
65. When the teacher says that she is going to give the class a test, do you get a nervous or funny feeling?	Yes No
66. While you are taking a test do you usually think you are doing poorly?	Yes No
67. While you are on your way to school, do you sometimes worry that the teacher may give the class a test?	Yes No

Appendix I:

Survey Directions

Directions read by examiner prior to survey administration*

My name is _____. I'm here from Michigan State University because I am interested in learning about children's feelings about tests. To do this, we are asking you and other children to take part in a research study. In this study, we will be asking you some questions about your feelings and behaviors.

If you agree to be part of the study, you will be asked to answer some questions that will take about 20 minutes and then we will give you _____ as a thank you! You do not have to answer any of the questions if you do not want to or if they make you feel uncomfortable.

No one will know the information belongs to you and the results of the research will not be able to be connected to you. Only information that you tell us that we think is dangerous to your health or others will be shared with your parents or others who can help keep you or others safe.

If you are interested in helping us learn how students feel about tests today, please write your name at the bottom of the page and write the date (write the date on the chalkboard for students to copy). Also, please write the name of your school, teacher's name, and grade on this page (write info on chalkboard for students to copy). **Give students time to finish reading the assent form you just summarized and to fill out the first page.**

The questions I'm going to ask you are different from the questions you are normally asked in school because there are no right or wrong answers. The questions are about how you think and feel, and therefore, they have *no* right or wrong answers. People think and feel differently, so the person sitting next to you may put down one answer and

you might put down another answer. For example, if I asked you this question: "Do you like to play ball?" some of you would answer "yes" and some of you would answer "no." Your answer depends on how *you* think and feel.

These questions are about how you thought and felt while you were taking [name test condition]. For these questions, think about how felt while you took [name test condition]. You should listen to each question as I read it aloud and then put a circle around your answer. Remember, listen carefully to each question and answer it by deciding how you think and feel. If you don't understand a question, raise your hand and ask me about it.

Adapt order of next paragraphs based on the order of questionnaires

BASC-2-TA

For the first few questions, I will say a statement and you have to decide if you agree or disagree. If you agree, circle *T for True.* If you disagree, circle *F for False.* ***Read Questions 1-3.***

For the next four questions, I will say a statement and you have to decide if you feel that way *never*, *sometimes*, *often*, or *almost always*. Circle 1 if you want to answer *never*, 2 if you want to answer *sometimes*, 3 if you want to answer *often,* and 4 if you want to answer *almost always*. ***Read Questions 4-7.***

CTAS

For the next set of questions, I will say a statement and you have to decide if you feel that way *almost never*, *some of the time*, *most of the time*, or *almost always*. Circle 1 if you want to answer *almost never*, 2 if you want to answer *some of the time*, 3 if you

want to answer *most of the time,* and 4 if you want to answer *almost always*. Now, turn the page to here's the first statement. Question 1: _____.... ***Read Questions 1-30 and add the "While I am taking [test condition]" starter to each statement. Remind students to turn the page after #15.***

TASC

For the last set of questions, I will say a statement and you have to decide if you agree or disagree. If you agree, circle *Yes.* If you disagree, circle *No.* ***Read Questions 1-30.***

*Directions adapted from Sarason et al., 1960.

REFERENCES

Abrams, L. M., Pedulla, J. J., & Madaus, G. F. (2003). Views from the classroom: Teachers' opinions of statewide testing programs. *Theory into Practice, 42*, 18-29.

American Educational Research Association. (2000). AERA position statement concerning high-stakes testing in PreK-12 education. Retrieved November 18, 2007 from http://www.aera.net/about/policy/stakes.htm.

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders, 4th edition, Text revision.* Washington, D.C.: American Psychological Association.

Barksdale-Ladd, M. A., & Thomas, K. F. (2000). What's at stake in high-stakes testing: Teachers and parents speak out. *Journal of Teacher Education, 51*, 384-397.

Bednarek, D. M. (1985). A study of test anxiety and helplessness-mastery orientation under experimental conditions of success and failure (Doctoral dissertation, University of Pittsburgh, 1984*). Dissertation Abstracts International, 46,* DA8525835.

Beidel, D. C., Turner, M. W., & Trager, K. N. (1994). Test anxiety and childhood anxiety disorders in African American and White school children. *Journal of Anxiety Disorders, 8,* 169-179.

Beidel, D. C., & Turner, S. M. (1988). Comorbidity of test anxiety and other anxiety disorders in children. *Journal of Abnormal Child Psychology, 16*, 275-287.

Beidel, D. C., Turner, S. M., & Taylor-Ferreira, J. C. (1999). Teaching study skills and test-taking strategies to elementary school students. *Behavior Modification, 23,* 630-646.

Bodas, J., & Ollendick, T. H. (2005). Test anxiety: A cross-cultural perspective. *Clinical Child and Family Psychology Review, 8*, 65-88.

Brown, D., Galassi, J. P., & Akos, P. (2004). School counselors' perceptions of the impact of high-stakes testing. *Professional School Counseling, 8*, 31-39.

Cheney, D., Flower, A., & Templeton, T. (2008). Applying Response to Intevention metrics in the social domain for students at risk of developing emotional or behavioral disorders. *The Journal of Special Education, 42*, 108-126.

Cizek, G. J. (2001) More unintended consequences of high-stakes testing. *Educational Measurement: Issues and Practice, 20,* 19-27.

Cizek, G. J., & Burg, S. S. (2006). Addressing test anxiety in a high-stakes environment: Strategies for classrooms and schools. Thousand Oaks, CA: Corwin Press.

Ergene, T. (2003). Effective interventions on test anxiety reduction. *School Psychology International, 24*, 313-328.

Fielding, C. (2004). Low performance on high-stakes test drives special education referrals: A Texas survey. *The Education Forum, 68*, 126-132.

Griffiths, J., & Peterman, S. (2007). MEAP Reporting. Retrieved February 25, 2008 from http://www.michigan.gov/documents/mde/23_MEAP_Reports_ 218829_7.ppt.

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

Hill, K. T. & Sarason, S. B. (1966). The relation of test anxiety and defensiveness to test and school performance over the elementary-school years: A further longitudinal study. *Monographs of the Society for Research in Child Development, 31*(2), 1-76.

Hill, K. T. & Wigfield, A. (1984). Test anxiety: A major educational problem and what can be done about it. *The Elementary School Journal, 85*, 105-126.

Hollingshead, A. B. (1975). Four factor index of social status. New Haven (Connecticut): Department of Sociology, Yale University.

Holroyd, K., Westbrook, T., Wolf, M,, & Badhorn, E. (1978). Performance, cognition, and physiological responding in test anxiety. *Journal of Abnormal Psychology, 87*, 442-451.

Hopkins, K. D., & Gullickson, A. R. (1992). Response rates in survey research: A meta-analysis of the effects of monetary gratuities. *Journal of Experimental Education, 61*, 52-62.

Jones, B. D., & Egley, R. J. (2004, August 9). Voices from the frontlines: Teachers' perceptions of high-stakes testing. *Education Policy Analysis Archives, 12(39)*. Retrieved November 12, 2007 from http://epaa.asu.edu/epaa/v12n39/.

Jones, B. D., & Egley, R. J. (2006). Looking through different lenses: Teachers' and administrators' views of accountability. *Phi Delta Kappan, 87*, 767-771.

Jones, M. G., Jones, B. D., Hardin, B., Chapman, L., Yarbrough, T., & Davis, M. (1999). The impact of high-stakes testing on teachers and students in North Carolina. *Phi Delta Kappan, 81*, 199-203.

King, N. J., & Ollendick, T. H. (1989). Children's anxiety and phobic disorders in school settings: Classification, assessment, and intervention issues. *Review of Educational Research, 59*, 431-470.

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

Lowe, P. A., Lee, S. W., Witteborg, K. M., Prichard, K. W., Luhr, M. E., Cullinan, C. M., et al. (2008). The Test Anxiety Inventory for Children and Adolescents (TAICA): Examination of the psychometric properties of a new multidimensional measure of test anxiety among elementary and secondary school students. *Journal of Psychoeducational Assessment, 26, 215-230.*

Matters, G., & Burnett, P. C. (2003). Psychological predictors of the propensity to omit short-response items on a high-stakes achievement test. *Educational and Psychological Measurement, 63*, 239-256.

McDonald, A. S. (2001). The prevalence and effects of test anxiety in school children. *Educational Psychology, 21*, 89-101.

Meijer, J. (2001). Learning potential and anxious tendency: Test anxiety as a bias factor in educational testing. *Anxiety, Stress, and Coping, 14*, 337-362.

Mulvenon, S. W., Connors, J.V., & Lenares, D. (2001, November). Impact of accountability and school testing on students: Is there evidence of anxiety? Paper presentation at the 30th Annual Meeting of the Mid-South Educational Research Association, Little Rock, AR.

Mulvenon, S., Stegman, C., & Ritter, G. (2005). Test Anxiety: A multifaceted study on the perceptions of teachers, principals, counselors, students, and parents. *International Journal of Testing, 5*, 37-61.

National Association of School Psychologists. (2003). Position statement on using large scale assessment. Retrieved November 18, 2007 from http://www.nasponline.org /about_nasp/ pospaper_largescale.aspx.

National Commission on Excellence in Education. (1983). *A nation at risk: The imperative for education reform.* Washington, DC: U.S. Department of Education.

Pallant, J. (2007). *SPSS Survival Manual: A step by step guide to data analysis using SPSS for Windows, third edition.* Berkshire, England: Open University Press.

Putwain, D. W. (2007). Test anxiety in UK schoolchildren: Prevalence and demographic patterns. *British Journal of Educational Psychology, 77*, 579-593.

Putwain, D. W. (2008). Do examinations stakes moderate the test anxiety-examination performance relationship? *Educational Psychology, 28*, 109-118.

Reynolds, C. R., & Kamphaus, R. W. (2004). *BASC-2: Behavior assessment system for children, second edition manual.* Circle Pines, MN: AGS Publishing.

Ryan, K. E., & Ryan, A. M. (2005). Psychological processes underlying stereotype threat and standardized math test performance. *Educational Psychologist, 40*, 53-63.

Salmon, G., & Kirby, A. (2008). Schools: Central to providing comprehensive CAMH services in the future. *Child and Adolescent Mental Health, 13*, 107-114.

Sarason, S. B. (1959). What research says about test anxiety in elementary school children. *NEA Journal, 48*, 26-27.

Sarason, S. B., Davidson, K. S., Lighthall, F.F., Waite, R., & Ruebush, B. K. (1960). *Anxiety in elementary school children.* New York: Wiley.

Sarason S. B., Hill, K. T., & Zimbardo, P, G. (1964). A longitudinal study of the relation of test anxiety to performance on intelligence and achievement tests. *Monographs of the Society for Research in Child Development, 29*(7), 1-51.

Seipp, B. (1991). Anxiety and academic performance: A meta-analysis of findings. *Anxiety Research, 4*, 27-41.

Smith, T. W., Ingram, R. E., & Brehm, S. S. (1983). Social anxiety, anxious preoccupation and recall of self-relevant information. *Journal of Personality and Social Psychology, 44*, 1276-1283.

Spielberger, C. D. (1966). The effects of anxiety on complex learning and academic achievement. In. C. D. Spielberger (Ed.). *Anxiety and behavior* (pp. 261-398). New York: Academic Press.

Spielberger, C. D., Anton, W. D., & Bedell, J. (1976). The nature and treatment of test anxiety. In M. Zuckerman & C. D. Spielberger (Eds.). *Emotions and anxiety: New Concepts, methods, and applications* (pp. 317-344). Hillsdale, NJ: Erlbaum.

State of Michigan. (2004a). Design and validity of the MEAP test. Retrieved February 25, 2008 from http://www.michigan.gov/mde/0,1607,7-140-22709_31168-94522--,00.html.

State of Michigan. (2004b). Requirements for schools not making AYP. Retrieved February 25, 2008 from http://www.michigan.gov/mde/0,1607,7-140-22709_22875-85932--,00.html.

State of Michigan. (2007). MEAP Fall 2007 Guide to Reports Grades 3-9. Retrieved April 27, 2008 from http://www.michigan.gov/documents/mde/8_Fall_2007_Guide_to_Reports_230806_7.pdf.

Steele, C. M. & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African-Americans. *Journal of Personality and Social Psychology 69*(5), 797-811.

Stöber, J., & Pekrun, R. (2004). Advances in test anxiety research. *Anxiety, Stress, and Coping, 17,* 205-211.

Triplett, C. F., & Barksdale, M. A. (2005). Third through sixth graders' perceptions of high-stakes testing. *Journal of Literacy Research, 37*, 237-260.

Tryon, G. S. (1980). The measurement and treatment of test anxiety. *Review of Educational Research, 50*, 343-372.

Turner, B. G., Beidel, D. C., Hughes, S., & Turner, M. W. (1993). Test anxiety in African American school children. *School Psychology Quarterly, 8*, 140-152.

U.S. Congress. (1994a). Public Law 103-227. *Goals 2000: Educate America Act.* Washington: Author.

U.S. Congress. (1994b). Public Law 103-382. *Improving America's Schools Act of 1994.* Washington: Author.

U.S. Congress. (2002). Public Law 107-110. *No Child Left Behind Act of 2001.* Washington: Author.

Wigfield, A., & Eccles, J. S. (1989). Test anxiety in elementary and secondary school students. *Educational Psychologist, 24*, 159-183.

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

Wong, K. K., & Nicotera, A. (2007). *Successful schools and educational accountability: Concepts and skills to meet leadership challenges.* Boston: Pearson Education, Inc.

Wren, D. G., & Benson, J. (2004). Measuring test anxiety in children: Scale development and internal construct validation. *Anxiety, Stress, and Coping, 17*, 227-240.

Yerkes, R. M., & Dodson, J. D. (1908). The relation of strength stimulus to rapidity of habit-formation. *Journal of Comparative and Neuropsychological Psychology, 18*, 459-482.

Zeidner, M. (1997). *Test anxiety: A transactional perspective*. Occasional Paper. Laboratory for Cross-Cultural Research in Personality and Individual Differences, University of Haifa, Haifa, Israel.

Zeidner, M. (1998). *Test Anxiety: The state of the art*. New York: Plenum Press.