Increasing Stakes in Testing:

The Impact on Test Anxiety in Middle School Students

Matthew Samelian

University of Wisconsin- River Falls

Author's Note

Matthew Samelian, Department of Counseling and School Psychology, University of Wisconsin-River Falls.

Correspondence concerning this article should be addressed to Matthew Samelian, Department of School Psychology, 257 Wyman Education Building, 410 South 3rd St, University of Wisconsin-River Falls, River Falls, WI 54022

E-mail: Matthew.Samelian@uwrf.edu

Abstract

Over the past 40 plus years, research on the construct of test anxiety has resulted in multiple theoretical models and much sociological and psychological data. The study being presented examined whether different types of assessments elicited different levels of anxiety in students. Two types of assessments, a high stakes test and a curriculum based exam, were compared to a baseline measure of state anxiety to determine the impact of exam stakes on anxiety. The State-Trait Anxiety Inventory was administered to a group of middle school students in a Midwestern school during the spring trimester. Results indicated that students experience a wide range of state anxiety related to the school experience, but do not show an increase in mean anxiety scores according to the stakes of the exam when compared to the baseline state anxiety measure.

Increasing Stakes in Testing: The Impact on Test Anxiety in Middle School Students

Examination stakes in our educational system are rising. The *No Child Left Behind* Legislation has aimed to simultaneously require local school districts to annually measure student progress via standardized testing as well as hold schools accountable for not making adequate yearly progress toward improving scores on these tests. School funding as well as reputation is at stake yearly as the results are made public and individual buildings are critiqued. Because of this, it appears education may becoming less an art-form of blending student interests and curriculum and more an exercise in uniformity and performance as evidenced by the performance-outcome monitoring associated with *No Child Left Behind*. The pressure to perform has been felt by all stakeholders in education and the frequency and intensity of exams our students' must take appears to be increasing. Within this context, educators may need to increase their consideration of what sorts of extraneous variables may influence student performance. All educational stakeholders have an obligation to fully understand the repercussions of the performance culture being created.

Casbarro (2004) succinctly described the undesired consequences that test anxiety and high stakes testing can have on teachers, students, administrators and communities. He reported instances of elementary students being unable to sleep or vomiting before a test and teachers who were disinclined to teach certain grades where high stakes test results are used for "school to school comparisons." An administrator's job performance may be judged on test results, and finally test outcomes can influence local property values if families flock to the highest performing schools and away from low performing schools.

Additionally, Black (2005) highlighted the recent increase in attention to the problem of test anxiety citing “greater pressure from teachers and parents to succeed, more testing in early grades, and the high stakes (such as retention and graduation) associated with many standardized tests” (Black, 2005). It appears that the stakes of performance based tests have been going up for school districts as well as students and communities as a whole with funding or restructuring of individual schools at stake based on terms of the *No Child Left Behind* legislation.

With these issues in mind, Gregor (2005) conducted a study to determine if school personnel could impact the severity of test anxiety by giving students interventions pertaining to management of anxiety. Results of the study demonstrated a difference in student ability to manage exam anxiety as measured by teacher ratings, self-ratings and in exam performance in Math when comparing the intervention group to the control group. Knowing that educators can impact the outcome of test anxiety with interventions requires one to take the next step in understanding when and where educators should be focusing intervention efforts.

These issues serve as the basis for the present study which aimed to examine whether different types of exams elicit different levels of anxiety specifically in middle school students. The results of this study will help educational stakeholders appraise the validity of using results from high stakes testing to make important educational decisions.

The construct of Test Anxiety and Theories

The literature examining the idea of test anxiety is exhaustive with researchers acknowledging the existence of over 1000 publications on the subject since the early 1950’s (Putwain, 2007). Research on test anxiety has been published in a wide variety of journals ranging from psychological journals to medical journals, to educational journals. This is a testament to the wide array of stakeholders in our educational system. The sheer volume of

research demonstrates a consensus that the construct of test anxiety is a serious issue. Through the years researchers have determined there are two factors present in the construct of test anxiety. These are the factors of worry and emotionality (Spiegler, Morris, & Liebert, 1968).

Worry, according to Spiegler, Morris and Liebert, pertains to thoughts and concerns about one's performance (1968). After administering a questionnaire designed to measure test anxiety Mandler and Sarason found that "In general . . . subjects with a high score on a test-anxiety questionnaire performed significantly worse than those with low scores. This effect has been found to hold only under ego involving instructions" (1968, p. 336). This suggests that a key component of the test anxiety construct is a perceived threat to self-worth. If students do not believe their performance on a particular test is important than they will most likely not be affected by the worry component.

The second component of the test anxiety construct that is described by researchers is considered the "emotionality" construct. "The emotionality dimension refers to the physiological component of anxiety, indicating high arousal of the autonomous nervous system and an unpleasant affective state characterized by nervousness and tension" (Friedman & Bendas-Jacob, 1997, p. 1036). Spiegler, Morris and Liebert found the emotionality construct of test anxiety was the result of the immediate testing conditions and environment and tended to increase and subsequently decrease post-test (1968, p.454).

Research on Demographic Differences

Putwain notes that the majority of research on the construct of test anxiety has been conducted in North America (2007). However, his research and the research of Zeidner and Safir have brought international attention to the issue of test anxiety. A primary goal of these studies

was to examine the differences in test anxiety level among different groups within the school population.

Zeidner and Safir specifically sought to measure the effects of “sex, ethnicity and social class on levels of test anxiety” (1989). Their results showed a significantly higher score for females than males, with sex accounting for 4% of the variance in scores. They also found this trend of higher scores for females to be true across ethnic groups (1989). They also found a significant difference in anxiety levels between middle and lower class students accounting for 6% of the total variance with lower class students reporting higher levels of anxiety. Despite this, Zeidner and Safir concluded the impact that test anxiety had on performance was small noting a weak correlation coefficient of -.16 between anxiety score and GPA and the small amount of variance explained by demographic differences (1989). Paradoxically, they also wrote that “although the correlations between test anxiety and achievement are low, there was an inverse relationship between test anxiety and test scores for girls, suggesting that test anxiety may be more debilitating for girls” (1989, p. 183).

The idea of a gender difference in test anxiety between boys and girls appears to be a significant one. It raises an important question of defensiveness in response to questions of anxiety. A meta-analysis by Hembree (1988) showed that at all grade levels, females exhibit higher test anxiety than males but that higher test anxiety did not necessarily translate into performance difference. This might suggest that another variable such as defensiveness to report anxiety could account for the difference in gender scores on test anxiety measures. It is important to note that Hembree’s meta-analysis showed a significant relationship between test anxiety and performance from grade three and up.

Additional research on demographic differences in the UK by Putwain showed higher self-reporting of anxiety by females than males once again. Counter to Zeidner and Safir, he noted that students from higher economic classes demonstrated slightly higher levels of test anxiety in the UK and that students from Black, Asian, and other ethnic backgrounds had higher anxiety scores than white students (2007).

In the United States Turner, Beidel, Hughes and Turner (1993) conducted a study examining test anxiety among African American students. Their goal was to identify racial differences in prevalence rates and their study consisted of 168 students in urban Pittsburgh schools. The results of their study found the prevalence rate of test anxiety to be 41% in their African American Sample. They compared this to prevalence rates from two other Pittsburgh area school districts and found 34% in a predominantly white middle class school district and 36% in a mixed racial urban school district. Their conclusion was that there was not a statistically different prevalence rate among three schools with which they had data. Although, the prevalence rates were not statistically different, the authors described their results as alarming.

Accurately measuring the prevalence rate of test anxiety has been a consistent problem noted throughout this literature review by differences in prevalence rates and different instruments and different testing scenarios. McDonald (2001) reported that the Test Anxiety Scale for Children was the most frequently used measure but lacked adequate norms. Other test anxiety specific measures that have been used are the Test Anxiety Inventory (Putwain, 2007), the Westside Test Anxiety Scale (Driscoll, 2007), the Test Anxiety Inventory for Children and Adolescents (Whitaker Sena, Lowe, & Lee, 2007), and the Test Anxiety Questionnaire (Mandler

& Sarason, 1968). Prevalence rates obtained from previous studies have ranged from 10% to 25-30% (McDonald, 2001, 93).

Research on Exam Stakes

A goal of this research was to find out whether the construct of test anxiety had a disparate impact on student performance on high stakes exams. Another goal was to expand understanding of the construct of test anxiety as it relates to different types of assessments. The earliest research available on exam stakes modifying test anxiety scores can be found in literature discussing a transactional model of test anxiety. Sawyer and Hollis-Sawyer studied three models of test anxiety construct, cognitive appraisal model, personality trait model and transaction model (2005). Cognitive appraisal model is reported to consist of different components such as evaluation of threat to self and the social surroundings of the test while personality trait model suggests that the major personality traits will account for most of the variance in test anxiety scores. The combination of these two has been termed as a transactional model. Results from their study suggest that most of the variance in the test anxiety construct can be explained best by the interplay of personality traits and cognitive evaluation of the immediate testing environment or assessment of the environment (Sawyer and Hollis-Sawyer, 2005). From this, educators can conclude the context surrounding the test is important, but it does not tell how or what testing scenarios are most likely to induce test anxiety.

Putwain conducted research in the UK asking the question of whether exam stakes moderate the anxiety/performance relationship. His work most closely resembles what is being considered in this study and in many served as a model for this study. Putwain was specifically

looking at students in their final two years of compulsory schooling which he describes as being "of critical importance to the future life trajectory of the student" (2008). He describes a situation in the student's academic career where all areas of study involve a "terminal examination" contributing either 40% or 50% of the students' total mark (2008). The stakes of these exams could be perceived as being as high as or higher than the tests given to students in the US. Putwain also noted that not all courses of study a student could choose to pursue had the same percentage of their mark decided by the high stakes terminal exam.

Contrary to his predictions, Putwain did not find that the higher stakes exams produced the most anxiety; instead he found the exam with the lowest stakes produced the highest anxiety. A key issue with his study was that each school permitted only one test session (2008). Because of this, the researchers could not measure a change in anxiety levels of particular students over time. This is important because individual student's have different reactions to testing conditions. The truest measure of test anxiety will show a change from a baseline measure. Secondly, Putwain labeled a mock exam as his lowest stakes variable. It is possible that "mock" exams, which might include pre-tests or practice tests, could be highly anxiety producing for students if their perceptions of these exams were that current performance will predict future performance. In this situation the way in which the "mock" exams were presented to the students would be very important. It is unclear how these "mock" exams were presented to students in this study. Also, in Putwain's research all self-reported anxiety measures were administered after taking the exam in question. Student's taking mock exams may realize how much more they need to prepare which may have caused the higher self-reported anxiety levels after the mock exams in his study.

An alternative but similar research design would be to compare anxiety levels between different types of assessments and also factor in student perceptions of importance which would measure how seriously they took each exam and subsequently control for unexpected anxiety related to different exams. Additionally new research should measure reported anxiety levels at the time of assessment which should accurately represent the students' "state" anxiety level at that point in time and not ask subjects to be reflective. The results obtained from individual changes in anxiety level will be the truest representation of increases in state anxiety as a direct result of the testing environment (Skybo, 2007).

Purpose

The purpose of this study was to examine whether two different types of assessments, the Minnesota comprehensive assessment and a curriculum based exam, produce different levels of anxiety among students in a suburban Minnesota middle school. To accomplish this, four specific research questions were developed and are as follows: a) Is there a significant difference between mean state and trait baseline anxiety level as measured by the State-Trait Anxiety Inventory and anxiety level prior to taking the Minnesota Comprehensive Assessment?; b) Is there a significant difference between the mean student baseline state and trait anxiety as measured by the State-Trait Anxiety Inventory and student self-reported state and trait anxiety prior to taking a curriculum based exam?; c) Are student self-reported anxiety levels prior to taking the Minnesota Comprehensive Assessment different from mean anxiety levels of the standardization sample as presented in the State-Trait Anxiety Inventory interpretive manual? And d) Do self-reported anxiety levels differ based on gender and race/ethnicity.

Initial hypotheses based on research by Putwain (2008) and Zeidner and Safir (2001) were that statistical analysis will show a difference between mean baseline anxiety levels and mean

anxiety levels just prior to both curriculum based exams and a high stakes test. Mean anxiety levels prior to both curriculum and high stakes tests were hypothesized to be higher than the standardization sample mean. Finally, it was also hypothesized that demographic variables such as gender and race/ethnicity will explain some of the variance in self reported anxiety levels.

Method

The study was conducted with administrative permission at a suburban Minnesota middle school. The study utilized two self-report measures: the State-Trait Anxiety Inventory for Adults and a short 19-item survey designed by the author. In addition to administrative support all students were required to have parent permission to participate in the surveys because of the sensitive nature of the information being gathered. In arranging accommodations for research with school administrators a concern was raised regarding student perceptions of the importance of the high stakes tests being administered in middle school. The 19-item survey that was administered was developed in order to validate results and provide anecdotal evidence of student attitudes for the cooperating building administrators. This was necessitated by criticisms that were made of Putwain's' research in the UK which highlighted potential areas of weakness in which he was unable to verify whether or not student's actually perceived "mock" exams as non-threatening. In this case the opposite is necessary. We must be sure that students attached sufficient importance to the Minnesota Comprehensive Assessment in middle school to warrant it being labeled as "high-stakes."

Participants

The sample for this study consisted of 35 students at a suburban Minnesota Middle School. Attrition resulted in 28 students completing all surveys requested by the author. The

school is diverse both racially and economically with 68% of the student body self-labeling as Caucasian, 8% Hispanic, 11% Asian and 11% African American. Additionally administrators report approximately 30% students on free or reduced lunch although free and reduced lunch status was not available for participants in this study due to school district privacy regulations.

Table 1

Demographic Data

Race	Frequency	Percent	Gender	Frequency	Percent
African-American	6	21.4	Male	12	43
Asian	2	7.1	Female	16	57
Hispanic	2	7.1			
White	18	64.4			

Although this sample size was small, it was representative with available sample demographics included in Table 1.

During the administration of the State-Trait Anxiety Inventory before the curriculum-based exam, 170 additional students who were not participating in the research for this study were given an anonymous 19-item survey designed by the author. This additional survey was conducted to provide anecdotal evidence of student effort levels and attitudes toward testing in middle school and was created at the request of building staff and as a result of concerns raised by school administrators regarding student perceptions of high-stakes tests.

Apparatus

The two measures that were administered for this study were the State-Trait Anxiety Inventory for Adults and a survey designed by the author titled Locus of Control, Effort and Efficacy Scale. The State-Trait Anxiety Inventory for Adults was chosen because it was a normed and standardized measure of anxiety that "has been used extensively in research and clinical practice . . . with more than 2000 studies using the STAI" (Spielberger, 1983, pp. 6, 8). Additionally the author chose this instrument because of the desire to use an established state anxiety measure that would be self-administering and to use a measure that had previously been used extensively in the research community. The trait anxiety component was less a concern of the author but because research by the instrument designer Spielberger demonstrated a correlation between state and trait anxiety reporting, both were administered. Additionally, according to Spielberger, "the STAIAD has been useful with junior high school students" (1983, p. 7).

To parlay concerns over using a measure normed on high school students and not middle school students, some measures were taken to eliminate these concerns. To eliminate vocabulary issues that became apparent in the initial screening of the instrument with students, the instrument was read and explained to each student in small groups prior to the first administration. Words that were found to be problematic included adequate and strained. These words were taught to each student and were re-taught on subsequent administrations. Because the students took the same survey three times and were coached on potential vocabulary issues the administrations should be seen as reliable. Additionally, the administration manual reports that key words in most STAIAD items are at the sixth grade reading level or below.

Statistics on the reliability and validity of the instrument were published in the administration manual and demonstrated 30 day test-retest reliability scores on the Trait Anxiety

measure ranging from .71 to .75. Thirty day test-retest reliability on the State-Anxiety component ranged from .34 to .62. Test-retest reliability should be lower on the State-Anxiety measure because by design it is measuring something that is seen as transitory. Cronbach's Alpha scores for the State-Anxiety measure ranged from .86 for males to .94 for females. The Trait-Anxiety measure reports Alphas of .90 and .90 for males and females respectively.

The survey designed by the author was a 19- item true or false survey that was administered on a larger scale than the anxiety inventory. This survey was designed to measure students' effort toward school and the MCAs through questions such as "When taking the MCA's I tried hard" and "I took taking the MCA's seriously". Other questions asked about student effort on Math tests and contained reverse loaded questions meant to function as a lie test. Another purpose of the survey was to measure students perceptions of the importance of the MCA's with questions such as "I think the MCAs are more important than a math test" and "Results from the MCAs are more important than a regular test." An examination of internal reliability of the survey produced an Alpha of .98 and demonstrated excellent internal reliability.

Procedure

The State-Trait Anxiety Inventory for Adults was administered to the sample three times during the spring of the 2008-2009 school year. Students were assured of confidentiality and were assigned numbers which replaced their names after the collection of surveys. A baseline measure of anxiety was taken first with participating students being read the instructions and taught vocabulary items in small groups and then taken into the library to record their responses at individual tables. The survey was administered according to the standardized instructions indicated in the survey manual. After the initial administration, students were asked to sign up to take the survey before taking the Minnesota Comprehensive Assessment (MCA) about three

weeks after the baseline administration. School administrators would not allow the survey to be given in the testing rooms and as such the survey was administered just before school in an empty classroom. The administration manual posits that the instrument can be effective in assessing situations that recently happened or will happen.

Approximately three weeks after the MCA administration, the final administration was given immediately prior to a Math quiz. This administration was given in the testing situation immediately after the teacher announced that the survey would be followed by a quiz. This can be seen as an authentic testing situation and is also when the additional survey measuring effort and attitude towards test was given. This administration consisted of participating students receiving the State-Trait Anxiety Inventory while other students in the class were given the Locus of Control, Effort and Efficacy Scale as an alternative. This survey was anonymous and no demographic data was collected.

Results

The Importance of High Stakes Testing

To begin the discussion of the findings of this study, results of the survey asking students to self-report their perceptions and feelings toward high stakes testing will be analyzed. As stated, this survey was given as a direct result of concerns from administrators that middle school students do not view high-stakes testing as important. Frequency results from the survey on attitudes and feelings about the MCAs will show evidence of student beliefs and are included in Table 2.

Survey results show that 44% of student's self-reported experiencing anxiety while taking the MCAs. Additionally 44% of students responded positively to a question asking if taking the MCAs was more stressful than a regular test. Results on effort related questions

showed that 96% of student reported taking the MCAs seriously and 95% reported trying hard. Reverse worded questions showed 7% of students didn't care about how well they did on the MCAs meaning 93% did care about assessment results. Additionally 5% of respondents reported that they did not try their best. The sample size of this survey was 170 which were more than half of

Table 2

Survey Questions and Results

Questions	Percent True	Percent False
Taking the MCA's was stressful	44%	56%
I took the MCA's seriously	96%	4%
When taking the MCA's I tried hard	95%	5%
the MCA's are more important than a classroom test	70%	30%
I didn't care about how well I did on the MCAs	7%	93%
When taking the MCAs I did not try my best	5%	95%
Taking the MCAs was more stressful than a regular test	44%	56%

all 7th grade students in the school. Analysis of this data shows that the middle school students in this particular school self-reported that they viewed taking the MCAs as important. Therefore results from the Minnesota Comprehensive Assessment measure should be seen as a valid reflection of student response to a testing scenario they viewed as important and worthy of effort.

Descriptive Statistics from STAIAD baseline administration

Because data showed that most students viewed the MCAs as important or more than a curriculum based exam, it was clear that further analysis of data obtained from STAIAD administrations had sufficient validity to be analyzed. Descriptive statistics of the baseline administration are given below in Table 3. The range of standard scores was from 33 to 65 on the

Table 3

Descriptive Statistics from Baseline Administration

	Mean	Std. Deviation	Skewness	Kurtosis
Baseline State Standard Score	46.25	9.2	0.741	-0.768
Baseline State Percent Rank	40.32	30.363	0.692	-1.039
Baseline Trait Standard Score	47.29	10.911	1.183	1.227
Baseline Trait Percentile Rank	42	28.891	0.618	-0.825

state anxiety measure and 34-76 on the trait anxiety measure. The mean state anxiety standard score was 46.25 with the standardization sample having a mean of 50. This means that the sample state anxiety mean standard score was below the 50^{th} percentile with the mean percentile rank being 40.32. Mean trait anxiety scores were 47.29 which is also slightly below the standardization sample mean of 50. Additionally measures of skewness and kurtosis are all less than +/- 2 and can be seen as acceptable.

A one-way ANOVA was conducted after baseline data was gathered to test for differences in reported baseline anxiety based on race and yielded an F statistic of 5.378 and a p-value <.05. Post-hoc analysis of results utilizing the Bonferonni correction showed a significant

difference between African-American and white students state anxiety raw scores with the African-American students reporting higher levels of state-anxiety, but no other significant differences were found. A one-way ANOVA was conducted using gender as the factor in question did not show a statistical different at the .05 significance level with an *F* statistic of 2.918 and a significance level of .10. .

Table 4

Descriptive Statistics from Curriculum-Based Exam

	Mean	Std. Deviation	Skewness	Kurtosis
Curriculum State Standard Score	45.5	9.3	1.036	0.068
Curriculum State Percentile Rank	44.19	27.401	0.569	-0.806
Curriculum Trait Standard Score	46.8	11.45	1.145	0.529
Curriculum Trait Percentile Rank	39.69	31.483	0.734	-0.802

Descriptive Statistics from STAIAD curriculum-based administration

Table 4 shows mean self-reported anxiety scores for the curriculum-based exam. Results from the curriculum-based administration of the STAIAD show a slight decrease in reported anxiety level from the baseline administration on both state and trait measures although a comparison of means showed the difference was not statistically significant. A one-way ANOVA was again conducted to determine if any differences existed between demographic groups. This yielded an *F* statistic of .329 and was not significant. This analysis showed that there was very little difference between state anxiety levels on the curriculum based exam. A

one-way ANOVA examining gender differences resulted in no group differences based on gender with an *F* statistic of .092.

Descriptive Statistics from STAIAD MCA administration

Final administration of the anxiety measure prior to taking the Minnesota Comprehensive assessment resulted in the following descriptive statistics in Table 5. Descriptive statistics from the MCA administration show a mean state anxiety score of 46.39 and a mean trait anxiety score

Table 5

Descriptive Statistics from MCA administration

	Mean	Std. Deviation	Skewness	Kurtosis
MCA State Standard Scores	46.39	9.2	1.172	0.609
MCA state percentile rank	40.57	27.577	0.871	-0.443
MCA trait Standard Scores	46.61	10.6	1.195	0.496
MCA trait percent rank	38.82	30.982	0.877	-0.676

of 46.61. The MCA state anxiety standard score mean was slightly higher than the baseline state anxiety mean which was 46.39 compared to 46.25. A one way ANOVA examined differences between racial groups yielded and *F* statistic of 1.424 and was not significant at the .05 level. Post-hoc analysis using the Bonferonni correction showed that no significant differences existed between specific groups. A one-way ANOVA for gender yielded an *F* statistic of .730 and was also not significant at the .05 level.

Comparison of Mean Scores for all three STAIAD administrations

Only one significant different was found based on self-reported state anxiety scores and was between African-American and White students on the baseline measure. When state raw scores were converted to standard scores the mean differences by race were as follows: on the baseline measure white students had an average score of 42.5 to the African-American students' average score of 57.5. This different was statistically significant. On the MCA administration White students had an average standard score of 44 while African Americans had an average score of 51. This difference was not significantly different.

MCA achievement scores were recorded on a scale from 1-4. Each number corresponded to categorical classification of student performance with 1 being does not meet standards, 2 being partially meets standards, 3 being meets standards and 4 being exceeds standards. A one way ANOVA was used to determine if results from the MCAs indicated any differences between groups. This ANOVA yielded an F statistic of 2.144 and a P-value of .121 which was not significant at the .05 level. Post-hoc analysis of scores by race showed African-American students in the sample scored 1.073 points lower than their white peers in the sample. This was not statistically different given the sample size in question.

The results listed will be summarized in the specific context of each research question. Research questions one and two asked if there was a statistical difference between baseline, MCA and curriculum-based exam anxiety levels. The mean standard score for baseline state anxiety was 46.25. The mean state anxiety standard score for the curriculum based measure was 45.5 and the mean state anxiety standard score prior to the MCAs was 46.39. There was not a statistically significant difference between means and as a result the answer to both research questions one and two is that no difference was observed between sample mean anxiety levels prior to the different types of assessments chosen and the baseline anxiety level. Research

question three asked if there was a significant difference between mean state anxiety levels reported prior to taking the Minnesota Comprehensive Assessment and the normative sample. The mean state anxiety standard score prior to taking the MCAs was 46.39. The standard score has a population mean of 50. A one sample t-test was statistically significant ($p<.05$). However, this difference was not in the direction anticipated and showed that students in the sample were reporting less state anxiety than in the normative population. Finally, the fourth research question asked if there was a difference in self-reported anxiety scores based on gender or race/ethnicity. To answer this question multiple one-way ANOVAs were used looking at each administration of the State-Trait Anxiety Inventory separately. Only one ANOVA showed a statistically significant difference between groups and this was the baseline administration. It did show a different between African-American and White students. On the MCA administration mean African-American scores were higher than White students but this was not statistically significant.

Discussion

Results from this study indicated that students in this sample did not experience a significant increase in state-anxiety levels just prior to high-stakes testing or curriculum based testing as compared to baseline levels. However, this study did show a significant difference in reported state anxiety levels between African-American students and White students on the baseline measure. The difference in self-reported anxiety level between White and African-American students was troublesome and the elevated levels of African-American student anxiety may have resulted in a higher than expected state anxiety level for the baseline administration. Research from McDonald (2001) suggested that prevalence rates in the literature vary widely from 10% to 30%. Research from Turner, Beidel, Hughes and Turner (1993) reported prevalence

rates in the 30%-40% range. With these prevalence rates in mind a sample of only 28 might have as few as 3 to 11 students suffering from test anxiety. In this study, it appeared that four students saw significant increases in their state anxiety scores on the MCA's as compared to their baseline state anxiety score. This results in a prevalence rate of 14%. The four students that saw the expected increase in state anxiety levels were likely not enough to statistically alter the results for the group as a whole.

The study presented in this article was in large part modeled after the work of Putwain (2008). Both studies aimed to gauge the impact of exam stakes on student self-reported anxiety level. Major differences between this study and that of Putwain were that this study measured the same students for each level of exam conceptualized and also utilized a true baseline instead of a "mock" exam for the least threatening measure of state-anxiety. This study also asked participants to self-report state anxiety immediately prior to the testing situation instead of asking them to reflect on past experiences.

Limitations

It was the hope of the author that the small sample size of 28 would yield statistically significant results and warrant future inquiry. Results that have been discussed did not yield statistically significant results with the exception of baseline state anxiety standard scores. Larger sample sizes may have resulted in significant differences between groups on the MCA state anxiety standard score as well as MCA results. However with only 6 Africa-American students in the sample, it was difficult to obtain statistically significant differences. One could argue, however, that the differences reported are very strong support for future studies, even if they were not statistically significant on the MCA and curriculum state measures of anxiety. In addition to sample size another issue with study design was the method of obtaining self-reported

scores prior to taking the MCAs. Due to administrator concerns, participating students were asked to take the survey immediately prior to taking the MCAs but not in the authentic testing environment. Students were given 15 minutes to complete the survey right as they came to school but before all students were required to report to their testing rooms. This scenario was less than ideal and may not have accurately measured students feelings of anxiety as evidenced anecdotally by 44% of students reporting that taking the MCA's was stressful.

The sample was likely too small to show the expected increase in state anxiety levels on the MCA and was also too small to show a negative impact on performance associated with elevated levels on anxiety. This study, did however provide much evidence to support future research on anxiety differences between African-American and White students, and also provides support for future exploration of an anxiety-performance relationship.

Implications for future research.

Future research should focus on identifying students who demonstrate significantly higher levels of state anxiety specifically associated with high stakes testing. True state anxiety baselines should be used to determine actual impact rates and research should attempt to answer questions about the impact of test anxiety on student outcomes. Additionally it is possible that middle school students do not experience as much stress surrounding high stakes tests as high school students. The exam stakes are different when there are "real-life" consequences, according to Putwain (2008). Because there are real life consequences such as graduation involved at the high school level, it makes sense for more research to focus on that arena. The fact that this study showed a mean difference in reported anxiety levels between White and African-American students may be a source of future inquiry specifically related to high-stakes testing performance as well. Once again, the small sample size presented here makes it hard to

draw firm conclusions, but that may be an area of great interest particularly to urban stake-holders in districts with a higher percentage of minority students.

References

Black, S. (2005). Test Anxiety. *American School Board Journal*, 192(6), 42-45.

Casbarro, J. (May/June, 2004). Reducing Anxiety in the Era of High-Stakes Testing. *Principle*, 36-38.

Driscoll, R. (2007). Westside Anxiety Scale Validation. (Retrieved April 6, 2008, from Online Source, ERIC No. ED495968).

Gregor, A. (2005). Examination Anxiety: Live With It, Control It or Make It Work For You?. *School Psychology International*, 26(5), 617-635.

Friedman, I., & Bendas-Jacob, O. (1997). Measuring Perceived Test Anxiety in Adolescents: A Self Report Scale. *Educational and Psychological Measurement*, 57(6), 1035-1046.

McDonald, A. (2001). The Prevalence and Effects of Test Anxiety in School Children. *Educational Psychology*, 21(1), 89-101.

Putwain, D. (2007). Test Anxiety in UK School Children: Prevalence and Demographic Patterns. Brittish *Journal of Education Psychology*, 77, 579-593.

Putwain, D. (2008). Do examinations stakes moderate the test anxiety-examination performance relationship? Educational Psychology, 28(2), 109-118.

Sadker, D. & Zittleman, K. (June, 2004). Test Anxiety: Are Students Failing Tests- Or Are Tests Failing Students. *Phi Delta Kappan*, 740-751.

Sawyer, T. & Hollis-Sawyer, L. (2005). Predicting Stereotype Threat, Test Anxiety, and Cognitive Ability Test Performance: An Examination of Three Models. *International Journal of Testing*, 5(3), 225-246.

Skybo, T. & Buck, J. (2007). Stress and Coping Responses to Proficiency Testing in School-Age Children. *Pediatric Nursing*, 33(5).

Spielberger, C. (1983). State-Trait Anxiety Inventory for Adults: Sampler set (manual, test booklet, and scoring key). Consulting Psychologists Press, Inc. www.mindgarden.com

Turner, B., Beidel, D. C., Hughes, S., & Turner, M. W. (1993). Test Anxiety in African American School Children. *School Psychology Quarterly*, 8(2), 140-152.

Zeidner, M., Safir, M. (1989). Sex, Ethnic, and Social Differences in Test Anxiety Among Israeli Adolescents. *Journal of Genetic Psychology*, 150(2), 175-185.

Whitaker Sena, J., Lowe, P. & Lee, S. (2007). Significant Predictors of Test Anxiety Among Students With and Without Learning Disabilities. *Journal of Learning Disabilities*, 40(4), 360-376.