Barry University
Institutional Repository

Theses and Dissertations

2014

Implications of Using High-Stakes Testing for At-Risk Populations

Dasha M, Highsmith

This work is brought to you for free and open access by the Theses and Dissertations at Barry University Institutional Repository. It has been accepted for inclusion in open access Theses by an authorized administrator of Institutional Repository.

IMPLICATIONS OF USING HIGH-STAKES TESTING FOR AT-RISK POPULATIONS

BY

Dasha M. Highsmith

A Directed Research Project

Submitted to the Faculty of
Barry University in partial fulfillment
of the requirements for the degree of
Specialist in School Psychology
Miami Shores, Florida

March 2014

BARRY UNIVERSITY

A Directed Research Project submitted in partial fulfillment of
the requirements for the degree of
Specialist in School Psychology

Implications of Using High-Stakes Testing for At-Risk Populations.

BY

Dasha M. Highsmith

Approved By:

Emilie Ney, Ph.D., NCSP
Assistant Professor of School Psychology

Fay Roseman, Ph.D.
Assistant Professor of Counseling

Terry Piper, Ph.D. Dean
Adrian Dominican School of Education

Abstract

High-stakes achievement tests are increasingly being used nationwide in order to make decisions about students' educational futures. For example, in Florida, performance on the Florida Comprehensive Assessment Test (FCAT) is used to determine academic proficiency in grades three through ten and also help govern graduation requirements for high school seniors. However, questions have been raised as to the validity and fairness of the FCAT in relation to the heterogeneous groups it targets. The purpose of this study is to examine the relationship between low performing schools in the state of Florida and their socioeconomic trends. Archival data from the 2011-2012 school year was obtained from Florida's Department of Education public records regarding FCAT 2.0 grade points and Title 1 status for 226 schools. It was hypothesized that students in Title 1 schools would score significantly lower on the FCAT than students in non-Title 1 schools. This hypothesis was evaluated using a Welsh Analysis of Variance (Welsh-ANOVA). It was also hypothesized that schools which have the double risk factors of being both predominantly low SES and predominantly minority status will score lower than those with just low SES. A Welsh-ANOVA was also used to evaluate this hypothesis. Findings of the present study supported both hypotheses. Title 1 funded schools obtained significantly less FCAT points than non-Title 1 funded schools. Additionally, Title 1 funded institutions with heightened minority populations scored significantly lower than the Title 1 funded institutions with heightened levels of majority populations.

Implications of Using High-Stakes Testing for At-Risk Populations.

Socioeconomic status (SES) plays a key role in defining many aspects of a society, including its development and its functioning on an ongoing basis. On an individual level, SES impacts the opportunities that will be afforded to the members of a society as well as which barriers they are likely to face. For example, medical issues (Braveman, Cubbin, Egerter, Williams, & Pamuk, 2010), mental health concerns (McLeod & Kaiser, 2004), legal interactions (Aaltonen, Kivivuori, & Mrtikainen, 2011), and substance abuse (Pope, Wallhagen, & Davis, 2010) trends all vary according to SES, just to name a few. Among the many factors SES impacts is education, which is of particular concern due to the far reaching effects that it can have on a child's future. Not only does low-SES limit students' access to education, but it is also associated with lower levels of student achievement, higher dropout rates, and lower performance on standardized tests (Orlich & Gifford, 2006). This in turn impacts the opportunity for higher education and, ultimately, employment later in life (Frempong, Ma, & Mensah, 2012). Lack of opportunity for educational advancement, therefore, serves as an additional barrier that prevents class mobility and perpetuates the transmission of low-SES from one generation to another. In other words, the cycle of low-SES risk factors and outcomes continues and limits the ability for societal improvement among this population (Orlich & Gifford, 2006).

One aspect of education that currently has great potential to impact the future success of students is high-stakes testing. High-stakes tests (HST) are criterion-referenced assessments used to measure students' level of mastery of academic material and progress toward a set of academic standards. These tests are considered "high-stakes" because their results are associated with a broad range of outcomes for all stakeholders involved (i.e. learners, administrators, and educators). High-stakes tests are

currently being used to hold teachers and schools accountable for student performance and to determine grade promotion and retention as well as high school graduation for students (Darling-Hammond, 2007; Florida Department of Education, 2011).

With such important educational decisions riding on test results, any biases that may exist in these tests are likely to create unfair barriers to academic and professional success for students of certain sub-populations. Therefore, it is important to establish whether high-stakes tests present a fair measure of academic ability for all students. One group that deserves particular attention due to the multiple risk factors they face is students from a low SES background. Criterion-Referenced tests have demonstrated apparent variations in student performance due to the students' economic and cultural dynamics (Darling-Hammond, 2007). For example, several factors have been identified that contribute to differential performance among social classes, including instructional techniques, educational opportunities, and performance expectations. Research has suggested that inequalities exist in both the school personnel and the learning environments of schools within the low-SES communities. For example; educators in low-SES environments have been found to have less experience, endorsements, and certifications than their counterparts in higher SES areas. Similarly, lower-SES schools have tended to have less or inadequate resources and a larger percent or poor readers (Baker &Johnston, 2010). Therefore, since these high-stakes tests are often also used to detect students who are at-risk for retention, are not making grade-level gains, or are displaying difficulties in specific subjects (Jordan, 2010); any biases that may occur on these tests are likely to create a ripple effect, multiplying the risk factors faced by the victims of such bias.

The purpose of this study is to examine the equity of high-stakes testing, more specifically the FCAT, in relation to the cultural and ethnic differences of those whom it assesses. Implications for improvement can be elicited based on the results in order to ensure that assessment tools adequately take into account the ethnic, cultural, and economic diversity of the population. Incorporating socioeconomic differences into test formation can begin to assess student learning in an effective manner, improving not only reliability, but validity of test results as well.

Risks and Outcomes Associated with Low-SES

Throughout the past few years, researchers have established compelling associations between social class and both student achievement and cognitive abilities (Ready, 2010). For example, studies have found students from low-SES environments began school significantly behind that of their same aged/grade peers from social classes superior to theirs; and these deficits worsen as the years progress (Downey, von Hippel, & Broh, 2004). Additionally, children of low-SES family environments are more liable to transfer homes, schools, and locations than higher-SES children and families (de la Torre & Gwynne, 2009). This habit of relocating or transferring from place to place greatly affects students' abilities to perform in the educational setting due to the inconsistency of their home life (de la Torre & Gwynne, 2009). A study conducted in Washington State discovered a negative correlation between reading scores and free and reduced lunch percentages at several public schools (Abbott & Joireman, 2001). Additional findings from this study proposed that as poverty increases, failure on the Washington Assessment of Student Learning (WASL) increases. Furthermore, outcomes from a Denver study reinforced the notion that low-income children demonstrated inadequate achievement

levels within the classroom in addition to demonstrating a high probability of performing poorly in high-stakes tests (Gottlieb, 2002).

Disadvantages among low social classes are evident due largely in part to these youth being unequally ascribed to educational settings which provide subpar resources and experiences that are essential to learning and academic growth (Tach & Farkas, 2006). It was found that students of low-income schools typically experienced under/unequally-qualified faculty and educational resources (Fram, Miller-Cribbs, & Van Horn, 2007). Additionally, students from disadvantaged communities tend to face challenges outside of the school environment that impact their academic performance. For example, they have been found to be predisposed to limited cultural and educational opportunities, receive limited guidance, assistance (i.e. economic, instructional, and technical) and reassurance in their homes, and are less likely to attain early development and literacy skills as opposed to their middle and upper-class counterparts (Baker & Johnson, 2010). The Early Childhood Longitudinal Study (ECLS) found children in high-SES environments attained high performance levels beginning in kindergarten as well as a rapidly increasing reading growth and vocabulary per month (U.S. Department of Education, National Center for Education Statistics, 2004). Students raised in low-SES environments, on the other hand, appeared to receive inadequate economic and parental support from home which in turn decreases their chances for achieving at or above grade-level (Baker & Johnston, 2010). Additionally, the youth in these populations were more likely to be living in a single guardian household and have parents with limited or low levels of education, which tests to result in limited exposure to reading or math related activities (Baker & Johnston, 2010). Results of past studies also discovered children from high-

poverty communities receive different and inadequate information relating to the advantages of education than children from low-poverty areas (Rouse & Barrow, 2006).

Schools in low-SES communities frequently attempt to remediate their students' performance deficits by developing compensatory programs such as after school tutoring, Saturday school, holiday school, as well as extended day and year school (Baker & Johnston, 2010). Although commendable, this occurrence creates heightened pressure and challenge to an already demanding environment for administration, faculty, and students. Cunningham and Sanzo (2002) proposed that the United States government should modify laws governing testing procedures in an effort to account for the barriers faced by disadvantaged youth. They suggest that increased awareness of the academic, social, and environmental barriers affecting these diverse students can lead to a more proactive approach for implementing appropriate interventions for these populations (Cunningham & Sanzo, 2002). Such programs would in turn relieve school administrators from bearing the majority of the pressure in addressing the issues of their populations. For instance, if the diverse needs of this population were taken into account at the legislative level, the pressure on administrators would reduce so that they would have more time and resources to dedicate to meeting the needs within the school. Although high-stakes tests seek to foster accountability, some argue that it is not conducive to collegiality between teachers and administrators, which likely has implications for the quality of education that is provided in the school (Cunningham & Sanzo., 2002).

In response to research highlighting the impact that SES has on student achievement, the government allocated additional money to low-performing schools serving at-risk students calling this additional money Title I funding of Elementary and Secondary Education. The Title I Funding Act of 1965 was introduced in an effort to

provide students of these populations with additional resources to supplement their academic performance (Title I Funding Act of 1965). The premise behind the additional funding and resources was to attempt to allow schools in low-SES communities a fair access to education and educational opportunities. Not only does this money target educational curriculum, but also interventions, technologies, and additional faculty to improve the student-teacher ratios (Title I Funding Act of 1965). Although this Act provides much needed support, it appears to have limited influence on low-SES student achievement when compared to the schools not receiving Title I funds. While the provision of these funds addresses some of the risk factors impacting children from low SES, there are several factors which this funding cannot resolve. Achievement gaps between high and low SES populations can be traced, beyond the typical environmental factors, to parent and teacher expectation as well as the method in which educational information is obtained (Rouse & Borrow, 2006).

High-Stakes Testing & the Florida Comprehensive Assessment Test

In order to address the underachievement of low-SES youth and at-risk groups, the No Child Left Behind Act was established in 2002. The ultimate goal of NCLB is geared towards promoting increased performance with regards to reading and mathematics in an effort to reduce the achievement gap throughout the country and among the diverse cultures and subcultures therein (NCLB; Pub. L. 107-110, enacted January 8, 2002). Achievement gaps reflect the differences in performance among different populations. In this instance it has been noted in various research that there lies a major difference in achievement between low-SES populations and their higher-SES counterparts (Baker & Johnston, 2010; Duran, 2005). NCLB is based on a theory proposing that people are motivated to perform their best when there are measurable standards of performance and

well defined consequences tied to these performance levels. Therefore, standards of success have been outlined for students, educators, and educational institutions and penalties have been identified for those who do not meet specified standards (Lee & Reeves, 2012). Additionally, NCLB regulations propose the reduction of achievement gaps between historically "disadvantaged and special groups of students" (Duran, 2005).

One of the cornerstones of NCLB is the concept of accountability, whereby school teachers and administrators are evaluated according to their students' academic performance. Educational institutions are expected to use student high-stakes test outcomes for appraisal of whether or not students are making progress toward academic standards set by the state. This practice is known as Adequate Yearly Progress (AYP). NCLB sets a goal of 100% academic proficiency; in other words, 100% of a school's student population should be performing on grade level by 2014. Additionally, schools are required to provide "continuous and substantial improvement" during a "reasonable" timeframe, so that all learners in the state could achieve a proficient level of performance in conjunction with state standards in all subjects. This goal of proficiency serves as the measure which the state uses to identify schools needing improvement. If the educational institutions persistently fail to make "adequate progress," the state provides additional support to the schools by introducing more intense interventions to improve their performance (Florida Department of Education, 2011). It is believed that the results of the high-stakes tests can provide lawmakers with empirical data to determine if students are progressing toward the goal. Every year institutions are evaluated based on these results and how much they have or have not progressed towards the ultimate goal of 100% proficiency (Lee & Reeves, 2012). This expectation of proficiency is based on the theoretical framework of performance-driven accountability, as it seeks to advise,

influence, and reorient educational stakeholders toward the goal by deeming schools, teachers, and students responsible for the performance of their students through incentives and penalizations in the form of merit pay and/or sanctions (Center on Education Policy, 2007; Lee & Reeves, 2012). This translates to the notion that if educators were held liable for the failure or success of their students, and if their pay were dependent on these outcomes, they would be more inclined to increase or improve educational instructions. Additionally, schools who fail to meet AYP also run the risk of being required to take corrective action in the form of curriculum changes, faculty replacement, updating or revamping improvement plans, and/or facilitating extended school year or day for at-risk students (Center on Education Policy, 2007).

Through NCLB, accountability of the schools is determined by assessing student performance with the administration of standardized tests. Each state develops these high-stakes tests which are then approved by the United States federal government (Florida Department of Education, 2011). High-stakes testing is intended to encourage fair, appropriate evaluation of student learning. On these examinations, students are to achieve an aptitude level of a predetermined minimum score. Since high-stakes test scores are often used to determine key decisions for the institution and its students, institutions are expected to guarantee these assessments are highly correlated to their current curricula as well as safe guard against biases by ensuring these assessments offer an equitable and valid measure of student proficiency (Duran, 2005). High-stakes test guidelines place emphasis on identifying and understanding students' academic difficulties with the use of assessments measuring primary and secondary grade-level curricula. Further parameters of high-stakes assessments include the provision of fair and impartial experiences of quality instruction despite cultural, ethnic, socioeconomic status, and/or physical and

mental limitations (Duran, 2005). Given these conditions, it is expected that every student should receive at least a minimum passing score on these standardized tests.

There have been many studies addressing not only the advantages of high-stakes testing but the latent results of such tests as well (Penfield, 2010; American Educational Research Association [AERA], 1999). For example, some educators believe the use of high-stakes testing provides diagnostic information in regards to students' strengths and weakness, as well as areas that may need consideration for improvement (Phelps, 2003). Other supporters of high-stakes testing propose the accountability factor can be beneficial in promoting thoroughness among educators and administrators regarding at-risk students (Phelps, 2003; Phelps, 2005). On the other hand, it has been found that the issue of retention under high-stakes tests is seen as more controversial than that of teacher-initiated retention because of the testing's more linear disposition (Allensworth, 2005). This is because under teacher-initiated retention the decision is based on multiple factors such as grades, attendance, and behavior as well as the teachers' subjective perspectives as opposed to the regulations of high-stakes tests where the determination to retain students is grounded solely on the performance on the test and takes less of a holistic assessment of the student.

The Florida Comprehensive Assessment Test (FCAT) is a high stakes test which public school students in the state of Florida must pass not only for grade promotion but high school completion as well. In third grade, students who fail to make passing scores on the FCAT reading are retained in an effort to reintroduce the content of that particular grade. The premise of this is based on the idea that if students have not passed those sections, they have not achieved grade-level proficiency and will need intervention. Additionally, on the secondary level, each student is allotted five additional opportunities to

retake the assessment if they are unsuccessful on the first. If they are unsuccessful on those additional attempts, they are rewarded a certificate of completion instead of the traditional high school diploma, which may be detrimental towards the pursuit of higher education (Florida Department of Education, 2011). With all of these factors being tied to a single test of this magnitude, it becomes difficult to determine if students in low-SES environments are truly lacking academic proficiency or if this type of assessment is sensitive enough to this population.

The unintended effects associated with high-stakes tests can be illustrated from a real life scenario that occurred in Minnesota. In 2002, over 7,000 students were mistakenly notified of failure on the Minnesota Basic Standards Test in Mathematics and were in jeopardy of not meeting graduation requirements (Cornell, Krosnick, & Chang, 2006). This unfortunate scenario provided researchers with a unique lens into the emotional impact that is brought about by tying test scores to such severe consequences as high school graduation. Cornell and colleagues (2006) administered a sequence of questionnaires assessing student and parent reactions to this situation and found that about 4% of students recounted dropping out of school as a direct outcome of the misinformation. Although this was an unfortunate scoring mistake on the part of Minnesota Department of Education, it must be known that these students were already on the borderline of failing because the scoring mistakes made were on about two-three questions on the actual exam (Cornell et al., 2006). This situation is a shocking demonstration of the direct negative impact that high-stakes tests had on influencing students' educational decisions.

High-Stakes Test Bias

The ultimate goal of high-stakes tests is to address and assess effective instructional practices; however some have argued that this intent has been

counterproductive in regards to the benefit of the students (Clarke, Shore, Rhoades, Abrams, Miao, & Li, 2003). Studies have reported the implementation of high-stakes tests is associated with reduced instructional creativity, elevated pressure to engage in test preparation, limited content coverage and incompatible instructional/curriculum structure and rate (Clarke et al., 2003). While these instructional practices are conducive to heightened FCAT scores, they are likely to be detrimental to student learning. In addition, the occurrence of incongruent instructional trends among educational institutions and communities of various socioeconomic statuses tend to be one-sided in favor of upper-class communities (Tach & Farkas, 2006).

Concerns regarding the impact that the FCAT has on key decisions such as retention, promotion, school performance grades, and teacher pay have risen within the last several years (Penfield, 2010). These concerns seem to stem from questions about the level of fairness this assessment has, or doesn't have, with regards to the populations it seeks to assess. More specifically, there appears to be unequal advantages as to the level of instruction students in different communities receive (Penfield, 2010). On a general level, there is a breadth of research expressing the belief that high-stakes tests provide accurate and reliable measures of student learning; yet limited research demonstrates this notion to be true for measuring learning in ALL students. In a study conducted by Horn (2003) findings yielded that high test scores do not directly predict student learning. Research shows that students living in low-SES and poverty communities have a high probability of attending low-performing schools which in turn predicts FCAT failure in reading (Myers & Curtiss, 2003).

In an effort to ensure that educational assessments are being used in a way that is valid, fair, and reliable for all members of the population, several sets of guidelines and

standards have been created, including the *Code of Fair Testing Practices in Education* (CFTPE) and the *Standards for Educational and Psychological Testing* (SEPT) (Joint Committee on Testing Practices, 2004). Using these guidelines and standards as a basis, the National Research Council (NRC) (Heubert & Hauser, 1999) established a set of principles intended to be used in evaluating the fairness of high-stakes testing procedures. The NRC offers a 3-component set of criterion for appraising the appropriateness of testing policies. These three components include validity and reliability, attribution of cause, and effectiveness of treatment. Validity and reliability refers to the statistical adequacy of an assessment. Attribution of cause is the extent to which the students being assessed have suitable learning conditions and instruction to learn the prerequisite skills and information presented on the test. Effectiveness of treatment evaluates the extent to which the assessment is being used in an effective and beneficial manner. The Attribution of Cause component is particularly relevant to the topic at hand, as it addresses the unfair advantage different ethnicities and cultures have in relation to accessing education. The question of concern is whether or not students' performance is related to their knowledge or skill acquisition, sub-par instruction, and/or exclusionary factors such as language, disabilities, or cultural barriers (Penfield, 2010).

It has been documented that minority populations; or populations that do not include the dominant culture of Caucasian ethnicity overall tend to face more challenges in the educational system as well as on high-stakes testing procedures (Brown, 2005). When students possess cultural and linguistic differences from the mainstream environment, testing procedures have potential to become problematic. Reliability and validity of testing results are called into question due to diversity of the populations required to take the test. Past research has reported that students reared in a diverse community have significantly

poorer educational experiences, language development, historical background knowledge of the mainstream culture, and various acculturation experiences when compared to their mainstream majority counterparts (Brown, 2005). Due to these differences among others, students of ethnic minority backgrounds tend to have a fairly inconsistent pattern of performance on high-stakes tests (Brown, 2005). Brown (2005) explained this notion by stating the content of standardized tests are derived from the knowledge base of the mainstream culture; however when the students who take the tests are not of mainstream culture or ethnicity they fall behind and the achievement gap increases significantly over time. This poses the question of the equity and validity of high-stakes tests such as the FCAT.

In the educational field, it is required by these standards that high-stakes testing used to make critical decisions (promotion & graduation) be accompanied with proof that every student have proper access to educational experiences and opportunities that deliver relevant knowledge and skills that are to be assessed. Prior research has indicated that there is an immense achievement gap between students of a minority background versus their majority counterparts (Baker, Griffin, & Choi, 2008; Heubert, 2002/2003). Additionally, these findings propose these disadvantaged youth are greatly predisposed to lower passing grades and test scores. With that said, it would benefit policy makers and test designers to be conscious of the limitations present in at-risk populations and reconstruct their testing procedures, content, and decisions accordingly. The current study seeks to evaluate whether the FCAT exemplifies this notion of an unequal advantage among at-risk (low-SES) populations and violates the above mentioned standard for Fair and Appropriate test use.

Purpose and Hypotheses

The purpose of this current study is to examine whether or not high-stakes tests, more specifically the Florida Comprehensive Assessment Test (FCAT), provides a fair measure on which to base important decisions that impact students, teachers, and administrators in schools in low-SES communities. The present study seeks to discover the association between educational institutions with a high population of minority and low-income students with their FCAT performance scores in reading and mathematics. In the previous mentioned studies, it was found that poor FCAT performance was more associated with urban/Low-income students than that of suburban/upper-class students. Since the majority of inhabitants in the urban community are at-risk individuals consisting of African American, Latino American, and other minorities of lower socioeconomic status (SES), examination of whether or not this factor is equal among minority cultures of various ethnicities will be conducted. These students are at-risk for a variety of negative outcomes in addition to the FCAT, and since this is the circumstance, it can be subsumed that this presents notable barriers to achievement not only in education but in employment and advancement into higher education as well; ultimately creating a cyclical pattern in the communities involved.

Limited empirical literature is available on the fairness and equity of the FCAT as a basis for high stakes decisions such as grade retention and accountability. As stated above, concerns regarding fairness and equity arise when the FACT and other high-stakes tests are used to determine grade retention and accountability. This is largely due to its neglect of additional factors associated with student performance. Therefore, the current study will focus primarily on evaluating the FCAT, so that stakeholders can make informed decisions on how best to use FCAT results. It was hypothesized that schools receiving Title 1 funding received a significantly lower amount of FCAT points than schools not

receiving Title 1 funding. It was also hypothesized that schools having the double risk factors of being both predominantly low SES and predominantly minority status received lower FCAT points than those schools with just low SES.

Methods

Participants

Participants included 226 high schools from the state of Florida. One-Hundred and thirteen of the schools included those supported by Title 1 funding during the 2011-2012 academic year and 113 were not supported by this funding. Of the 113 schools receiving Title 1 funding, 65 were predominantly minority schools with at least 70% of the student population from minority backgrounds. Only schools serving grades nine through twelve were included in this study, in order to eliminate variability between schools that might impact scores. Additionally, charter schools and private institutions were excluded from this study in order to control for confounding variables that might impact the students attending those schools. FCAT points were collected from the Florida Department of Education's public online database for the 2011-2012 school year.

Instruments

Socioeconomic status. The variable of low-SES was defined by the Title I status of the school. A school is identified as eligible for Title I Funding by the state of Florida's Student Achievement and School Accountability Programs office when the poverty level is at or above 40%. The poverty level is determined by the free and reduced meal counts, government assistance eligibility, census data, and/or Medicaid eligibility.

The Florida Comprehensive Achievement Test (FCAT). The Florida Department of Education (FLDOE) is currently operating under the Next Generation Sunshine State Standards which were generated to provide standards for instruction and promote student

achievement. In alignment with similar standards, the Florida Comprehensive Achievement Test (FCAT) was produced in 1998 as a means of assessing implementation and proficiency of the standards. In 2011 the FCAT was updated to align with the new standards, and is now identified as FCAT 2.0. Therefore, for the basis of this study, the FCAT 2.0 and Next Generation Sunshine State Standards will be used. Passing scores on the FCAT 2.0 indicate students are making "adequate" progress and achieving on grade level curricula. The FCAT 2.0 is a criterion-referenced test given to all students attending public schools. The FCAT reading and math portions are first introduced to students in third grade and continue to grade ten. Additionally, the science portion is administered in grades five, eight, and eleven; while the FCAT Writes exam is given in grades four, eight, and ten. Scaled scores on the on the FCAT 2.0 reading portion span between 140 to 302 while math scores span between140 to 298. These scores are then analyzed and computed by a computer program which allow for comparisons between students and across grade levels. From this computerized summary achievement levels are generated ranging from 1-5; with 3-5 reflecting satisfactory to mastery or proficiency. Currently, FCAT 2.0 scores are not modified to account for SES of students (Baker & Johnson, 2010). In fact, the Department of Education proposes education institutions are required to teach all students, despite their socioeconomic status and that all students possess the capacity to learn and make "Adequate Yearly Progress". This study examined the FCAT 2.0 overall points received during the 2011-2012 school year. Statistical analysis from the Florida Department of Education reports high content validity for the FCAT 2.0 (2007). Additionally, the FCAT is considered reliable as a result of the Cronbach's alpha for reading (0.882) and math (0.925) (FLDOE, 2012). According these results, the FCAT is

viewed as technically sound and meets or exceeds the professional standards for standardized achievement tests.

FCAT points. Information regarding the calculation of school performance grades was obtained through the Miami-Dade County Public Schools' Assessment, Research, and Data Analysis database. These grades are used as a method as established in Senate Bill1008.34 (2011), for school grading in 2009-10 and thereafter. "Florida's high schools are graded using the FCAT and learning gains components described in the grading criteria for elementary and middle schools (worth 800 points), plus several non-FCAT based components that account for 70 percent of the high school grade (also worth 800 points)," totaling 1600 points (Florida Department of Education, 2011; p.2). Schools earn one point for every percent of students scoring in achievement levels three to five in reading, math, and science; as well as one point for each percentage of students scoring three and a half or more in writing on a scale of one to six. Additionally, high schools have an opportunity to earn additional points if a minimum of fifty percent of students retaking the assessments achieve passing scores and meet graduation requirements in both reading and math. For the basis of this study, the FCAT points of a maximum 800 will be used as indicators of FCAT success or failure. Included in this score are the schools' reading, math, writing, and science performance and gains.

Procedure

High school FCAT 2.0 data from the FLDOE online database was used to obtain data on the variables of this study. Schools were first selected based on their Title 1 status during the 2011-2012 school year. All schools receiving Title I funding were included and an equivalent number of schools that did not receive Title I funding was randomly selected using a randomization function in SPSS. For each school, FCAT points attained during that

year as well as the percentage of minority students in the school was also collected. Any school in which 70 percent or more of the population is of ethnic minorities (i.e. Latino, African American/Black, Asian, Hawaiian/Pacific Islander, American Indian/Alaska Native, or multi-racial) was identified as a predominantly minority school. Each of the 226 schools was de-identified and their names were replaced with a participant number in order to ensure anonymity.

Results

The results are reported based on each hypothesis. Hypothesis one indicated that when controlling for ethnic makeup, schools receiving Title 1 funding received a significantly lower amount of FCAT points than schools not receiving Title 1 funding.

For hypothesis one, the assumption of homogeneity of variances was violated, as assessed by Levene's Test of Homogeneity of Variance (p=.005). Additionally, the Shapiro-Wilk Test of Normality discovered there was significant violation of normal distribution (p=.000). However further analysis revealed that both groups had similar distributions, so this violation was not considered to be of concern. To account for the violation of the homogeneity of variance, a Welch ANOVA was computed to test for a difference between the FCAT points of schools receiving Title 1 funding versus that of schools not receiving Title 1 funding. A significant difference was found between the two groups ($F(1, 187.264)=31.007$, $p<.0005$), with Non-Title 1 schools (M=472.87, SD=76.052) receiving higher scores than Title 1 funded schools (M=425.96, SD=47.255).

Hypothesis two stated that schools having the double risk factors of being both predominantly low SES and predominantly minority status will receive lower FCAT points than those schools with just low SES. The Levene's Test of Homogeneity of Variance (p=.001) displayed a violation of homogeneity at a statically significant rate. Additionally,

the Shapiro-Wilk Test of Normality discovered a significant violation of normal distribution (p=.000), however because both groups had similar distributions, this violation was not considered to be of concern. In order to account for the violation of homogeneity of variance, a Welch ANOVA was used to test the hypothesis. A statistically significant difference was observed between the low SES, predominantly minority schools and schools that are low SES but predominantly majority ($F(1, 69.628)=6.934$, $p=0.010$) with the non-minority Title 1 funded schools ($M=440.23$, $SD=58.514$) receiving statistically higher FCAT points than the minority Title 1 funded schools ($M=415.43$, $SD=33.588$).

Discussion

The present study intended to provide insight into the implications of using high-stakes testing procedures for at-risk populations. The first hypothesis suggested that schools receiving Title 1 funding would perform poorly, or receive less FCAT 2.0 points, than their Non-Title 1 counterparts. Results indicated that Non-Title 1 funded schools did attain significantly more FCAT 2.0 points than Title 1 funded schools. Therefore it is expected that although Title 1 funded schools or low-SES schools are receiving supplemental money for additional educational resources, they may still possess inferior aspects of academic instruction and resources when compared to those higher-SES schools. In addition to the quality of their education itself, students in Title 1 funded schools are likely exposed to environmental, financial, and familial factors that further hinder their performance among testing measures. This finding also supports past research relating to the effects that poverty and low-SES have on student academic performance and achievement scores on standardized tests (Abbott & Joireman, 2001; Gottlieb, 2002). Additionally, findings such as this bring attention to other factors that affect performance and achievement of low-SES and/or high poverty institutions such as subpar and

inconsistent instruction and resources (Baker & Johnston, 2010). The observed difference may also reflect possible testing bias in the FCAT 2.0; however this conclusion cannot be confirmed in the current study due to a lack of control for environmental factors that may impact academic performance. Regardless, the differential performance among groups and the implications that are associated with high-stakes tests results may lead to differential academic opportunity for different subgroups of the population, resulting in bias. Nevertheless, the present study's findings along with past research demonstrate that students from low-SES tend to score lower on high-stakes testing than students from higher levels of SES.

The present study supported the hypothesis that Title 1 funded minority institutions would receive less FCAT 2.0 points than their non-minority Title 1 funded counterparts. This suggests that within a low-SES population, students from different ethnic backgrounds differ on their performance on high-stakes testing. Therefore it appeared that ethnicity played a significant role in determining performance on standardized measures if these ethnicities attend low-SES, high poverty, and Title 1 funded institutions. It appears that SES as well as ethnicity plays a significant role in impacting high-stakes test scores.

It can be assumed that poverty and/or low-SES play an important role on a child's ability to perform in academic settings and on standardized measures. It appears that SES affects members of various ethnicities inconsistently when it comes to academic performance. This may be due to their variable environmental characteristics, resources, and other risk factors that may influence them within the educational institution. As proposed by Cunningham and Sanzo (2002), the risk factors associated with low-SES and poverty should be considered by policy makers when creating high-stakes tests so that each SES population has an equal and fair opportunity for achievement.

The present study had several limitations that should be considered when interpreting the results. A relatively small sample size of schools met criteria for inclusion in this research study. Of the 500 secondary schools in the state of Florida, only 226 met criteria for inclusion in the study after removing combination middle and high schools, special education and alternative schools, as well as private and charter schools. All of these institutions are either governed by different legislative regulations or have different scoring methods that could have potentially skewed the overall analysis and outcome of this study.

Another limitation of the present study related to the demographics of the sample. This study utilized school level data from high schools in the state of Florida, as opposed to individual student data. The results may not be as sensitive as if they were attained from individual student data and their performance on the test. This method would be beneficial for future research as insight could be gained based on examining individual student factors of being from a specific SES and ethnicity.

The third limitation of this study related to the inability to control for environmental factors associated with the minority and low-SES populations. If these variables were able to be manipulated the equality of the sample population would have helped generate more reliable, homogenous, and normally distributed results. This could have resulted in a clearer depiction of the role SES plays on standardized testing performance.

The final limitation of the present study is the short scope of the study in regards to the use of only one school year. It would be interesting to examine if the discovered trends existed in previous years and versions of the FACT. The information regarding this particular year could not have been a common happening and future research should

account for this notion and look towards multiple years to gain a clearer insight into the findings of past research to determine if the trends are still prevalent.

When it comes to further examining this association, one suggestion for future research would be to include the special education institutions into the analysis to determine if at-risk students were able to perform at a greater rate when placed in a more contained and specialized environment. Relating to the limitations mentioned above, it would be beneficial if environmental factors were controlled for to create cohesiveness among the sample population. Although the present study provided interesting findings as these schools are graded and held accountable based on their overall performance, future research should look towards examining this trend on a more individualized level versus the use of a school data.

References

Aaltonen, M.; Kivivuori, J.; & Mrtikainen, P. (2011). Social determinants of crime in a welfare state: Do they still matter? *Acta Sociologica, 54*(2), 161-181. doi: 10.1177/0001699311402228

Abbott, M.L. & Joireman, J. (2001). *The relationships among achievement, low income, and ethnicity across six groups of Washington state students* (Report No.1). Lynwood, WA: Washington School Research Center. Retrieved September 10, 2011, from http://files.eric.ed.gov/fulltext/ED529417.pdf

Allensworth, E.M. (2005). Dropout rates after high-stakes testing in elementary school: A study of the contradictory effects of Chicago's efforts to end social promotion. *Educational Evaluation and Policy Analysis, 27*(4), 341-364. doi:10.3102/01623737027004341

American Educational Research Association, *American Psychological Association, & National Council on Measurement in Education. (1999)*. Standards for educational and psychological testing. Washington, DC: American Psychological Association. Doi: 10.1037/14047-013.

Baker, M. & Johnson, P. (2010). The impact of socioeconomic status on high stakes testing reexamined. *Journal of Instructional Psychology, 37*(3), 193-199.

Braveman, P.A.; Cubbin, C.; Egerter, S.; Williams, D.R.; & Pamuk, E. (2010). Socioeconomic disparities in health in the United States: What the pattern tells us. *American Journal of Public Health, 100*(1), 186-196. doi: 10.2105/AJPH.2009.166082

Brown, J.E. (2005). Highly qualified minority teachers: Do high-stakes teacher tests weed out those we need most? *International Journal of Educational Policy, Research, &*

Practice, 6(1), p.105-137.Retreived from http://files.eric.ed.gov/fulltext/EJ795154.pdf

Center on Education Policy. (2007). *States continue trend toward higher-level exit exams, more subjects tested.* Retrieved from http://www.cepdc.org/document/docWindow.cfm?fuseaction=document.viewDocument&documented=209&documentFormatId=3627

Clarke, M., Shore, A., Rhoades, K., Abrams, L., Miao, J., & Li, J. (2003). *Perceived effects of state-mandated testing programs on teaching and learning: Findings from interviews with educators in low-, medium-, and high-stakes states*. Chestnut Hill, MA: Center for the Study of Testing, Evaluation, and Educational Policy, Boston College. Retrieved from http://files.eric.ed.gov/fulltext/ED474867.pdf

Cornell, D.G.; Krosnick, J.A.; & Chang, L. (2006). Student reaction to being wrongly informed of failing a high-stakes test. *Educational Policy, 20*(5), 718-751. doi:10.1177/0895904805284123

Cunningham, W. G., & Sanzo, T. D. (2002). Is high-stakes testing harming lower socioeconomic status schools? National Association of Secondary School Principals, *86*(631), 62-75. doi:10.1177/019263650208663106

Darling-Hammond, L. (2007). Race, inequality and educational accountability: The irony of 'No Child Left Behind'. *Race, Ethnicity & Education, 10*(3), 245-260. doi:10.1080/13613320701503207

de la Torre, M. and Gwynne, J. (2009). Changing schools: A look at student mobility trends in Chicago public schools since 1995. *Chicago: Consortium on Chicago School Research.* Retrieved from http://ccsr.uchicago.edu/sites/default/files/publications/studentmobility-final.pdf

Downey, D. B., von Hippel, P. T., & Broh, B.A. (2004). "Are schools the great equalizer? Cognitive inequality during the summer months and the school year." *American Sociological Review 69*:613-35. Retrieved from http://asr.sagepub.com/content/69/5/613.pdf

Duran, A. (2005). Factors to consider when evaluating school accountability results. *The Journal of Law and Education, 34*(1), 73-100. Retrieved from http://www.ppmrn.net/storage/ppmrn/factors.pdf

Florida Department of Education. (2011). *Graduation requirements for Florida's statewide assessments. Florida End-Of-Course (EOC) Assessments*. Retrieved from http://fcat.fldoe.org/eoc/

Fram, S.M.; Miller-Cribbs, J.E.; Horn, V.; & Lee. (2007). Poverty, race, and the contexts of achievement: Examining educational experiences of children in the U.S. south. *Social Work, 52*(4), 309-319. doi: 10.1093/sw/52.4.309

Frempong, G., Ma, X., & Mensah, J. (2012). Access to postsecondary education: Can schools compensate for socioeconomic disadvantages? *Higher Education, 63, 19-32.* doi: 10.1007/s10734-011-9422-27/s10734-011-9422-2

Gottlieb, A. (2002). Economically segregated schools hurt poor kids, study shows. *The Term Paper, I* (2), 1-8. Retrieved from http://www.piton.org/Documents/term2.pdf

Heubert, J. P. (2002/2003). First, do no harm. *Educational Leadership, 60*(4), 26–30. Retrieved from http://www.ascd.org/publications/educational-leadership/dec02/vol60/num04/First,-Do-No-Harm.aspx

Heubert, J. P., & Hauser, R. M. (1999). *High stakes: Testing for tracking, promotion, and graduation.* Washington, DC: National Academy Press. Retrieved from http://www.nap.edu/openbook.php?isbn=0309062802&page=R1

Horn, C. (2003). High-Stakes Testing and Students: Stopping or Perpetuating a Cycle of Failure? *Theory Into Practice, 42*(1),20 A\. Retrieved March 14, 2009, from http://www.jstor.org/stable/1477316 doi:10.1207/s15430421tip4201_5 doi:10.1353/tip.2003.0009

Joint Committee on Testing Practices. (2004). *Code of fair testing practices in education.* Washington, DC: Author. Retrieved from http://www.apa.org/science/programs/testing/fair-code.aspx

Jordan, W. J. (2010). Defining equity: multiple perspectives to analyzing the performance of diverse learners. *Review of Research in Education, 34*(1), 142-178. Retrieved from EBSCO*host*. doi:10.3102/0091732X09352898

McLeod, J.D. & Kaiser, K. (2004). Childhood emotional and behavioral problems and educational attainment. *American Sociological Review, 69*, 636-658. doi: 10.1177/000312240406900502

Myers, M. A., & Curtiss, D. (2003). Failing the equity test. *Principal Leadership, 4*(2), 70–73. Retrieved from http://www.principals.org/portals/0/content/48739.pdf

No Child Left Behind Act of 2001, Pub. L. No. 107-110. Retrieved from http://www2.ed.gov/policy/elsec/leg/esea02/107-110.pdf

Orlich, D.C. & Gifford, G. (2006, October). Test scores, poverty, and ethnicity: The new American dilemma. Paper presented at the meeting of Phi Delta Kappa's "Summit on Public Education," Washington, D. C.

Penfield, R.D. (2010). Test-based grade retention: Does it stand up to professional standards for fair and appropriate test use? *Educational Researcher, 39*(2), 110-119. doi:10.3102/0013189X10363007

Phelps, R. P. (2003). *Kill the messenger: The war on standardized testing.* New Brunswick, NJ: Transaction Publishing. Retrieved from http://richardphelps.net/KTMreviews.htm

Phelps, R. P. (2005). *Persistently positive: Forty years of public opinion on standardized testing.* In R. P. Phelps (Ed.), Defending Standardized Testing (pp. 1-22). Mahwah, NJ: Lawrence Erlbaum.

Pope R.C.; Wallhagen M.; & Davis H. (2010). The social determinants of substance abuse in African American baby boomers: effects of family, media images, and environment. *Journal of Transcultural Nursing. 21*(3), 246-56.

Ready, D. D. (2010). Socioeconomic disadvantage, school attendance, and early cognitive development: The differential effects of school exposure. *Sociology of Education, 83*(4), 271-286. Retrieved from http://soe.sagepub.com/content/83/4/271.full.pdf+html

Rouse, C. E., & Barrow, L. (2006). U.S. elementary and secondary schools: Equalizing opportunity or replacing the status quo? *The Future of Children, 16*(2), 99-123. Retrieved March 14, 2009, from http://www.jstor.org/stable/3844793. doi:10.1353/foc.2006.0018

Tach, L. M. & Farkas, G. (2006). "Learning related behaviors, cognitive skills, and ability grouping when schooling begins." *Social Science Research 35*:1048-079. doi:10.1016/j.ssresearch.2005.08.001

Title I of the Elementary and Secondary Education Act of 1965, 20 U.S.C. 6301 et seq. Retrieved from http://www2.ed.gov/policy/elsec/leg/esea02/pg1.html

U.S. Department of Education, National Center for Education Statistics. (2004). ECLS-K Longitudinal Kindergarten-Third Grade Public-Use Data File (CD-ROM). (NCES 2004-089). Washington, DC: Author.