1-1-2009

Gender And Performance In The Mathematics' Sections Of The Illinois Standards Achievement Test And The Prairie State Achievement Exam

Lindsey Leaf
Eastern Illinois University

THESIS MAINTENANCE AND REPRODUCTION CERTIFICATE

TO: Graduate Degree Candidates (who have written formal theses)

SUBJECT: Permission to Reproduce Theses

The University Library is receiving a number of request from other institutions asking permission to reproduce dissertations for inclusion in their library holdings. Although no copyright laws are involved, we feel that professional courtesy demands that permission be obtained from the author before we allow these to be copied.

PLEASE SIGN ONE OF THE FOLLOWING STATEMENTS:

Booth Library of Eastern Illinois University has my permission to lend my thesis to a reputable college or university for the purpose of copying it for inclusion in that institution's library or research holdings.

[signature] 3/23/09

Author's Signature Date

I respectfully request Booth Library of Eastern Illinois University **NOT** allow my thesis to be reproduced because:

____

____

____

Author's Signature Date

This form must be submitted in duplicate.

Gender and Performance on the Mathematics' Sections of the Illinois Standards Achievement Test and the Prairie State Achievement Exam

BY

Lindsey Leaf

THESIS

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF

Specialist in School Psychology

IN THE GRADUATE SCHOOL, EASTERN ILLINOIS UNIVERSITY
CHARLESTON, ILLINOIS

2009
YEAR

I HEREBY RECOMMEND THIS THESIS BE ACCEPTED AS FULFILLING
THIS PART OF THE GRADUATE DEGREE CITED ABOVE

3-23-09
THESIS COMMITTEE CHAIR DATE

3/23/09
DEPARTMENT/SCHOOL CHAIR OR CHAIR'S DESIGNEE DATE

3/23/09
THESIS COMMITTEE MEMBER DATE

THESIS COMMITTEE MEMBER DATE

3/23/09
THESIS COMMITTEE MEMBER DATE

THESIS COMMITTEE MEMBER DATE

Running head: Mathematics Achievement and Gender

Gender and Performance on the Mathematics' Sections of the Illinois Standards Achievement Test and the Prairie State Achievement Exam

Lindsey Leaf

School Psychology Program

Eastern Illinois University

23 March 2009

Abstract

Despite criticisms about their inability to assess students' learning accurately, high stakes tests are currently used in the United States to measure student and teacher progress. Because of the need for schools to demonstrate students have achieved or maintained satisfactory progress, schools in Illinois regularly administer the Illinois Standards Achievement Test (ISAT) and Prairie State Achievement Exam (PSAE), despite limitations in the statistical and technical information available for both of these tests. This study examined possible gender differences on the mathematics portion of the ISAT and PSAE by gathering data from one school district in Illinois. Findings indicated that while there were no overall gender differences between boys and girls on the ISAT or the PSAE, there were more male than female students who scored at the lower end of the distribution when data from both tests were combined. Additionally, there was a greater percentage of female than male students who achieved academic success standards with the PSAE. Finally, grade level comparisons showed that students in lower grade levels taking the ISAT were more likely to be successful in meeting or exceeding standards than high school students taking the PSAE. Results from the present study suggest, therefore, that boys and students in high school are at risk for not making satisfactory progress in math achievement as currently measured by high stakes testing. Future research is needed to confirm these findings and to identify possible causes and successful interventions.

Table of Contents

List of Tables & Figures

Gender and Performance on the Mathematics' Sections of the Illinois Standards Achievement Test and the Prairie State Achievement Exam

Because reading skills have long been recognized as crucial for success in school as well as in life, reading has been the focus of most previous research related to academic subject areas. Research on mathematical achievement, however, is on the rise because of recognition of the need for math skills throughout one's life and because of increasing concern over poor math performance by some students on standardized tests of achievement. To date, several cognitive (such as information processing skills) and noncognitive (e.g., quality of instruction) factors have been described by researchers as influencing the mathematics achievement of children and adolescents. Some of these factors may affect achievement throughout the school years, from kindergarten through twelfth grade, while others only affect mathematics achievement for a specific or short period of time. Additionally, some factors affecting mathematics achievement may be influenced by the child's home or school environment, while other factors may be genetic or biological in origin. For instance, there have been ongoing discussions and research questioning the role of gender on mathematics achievement. While some research findings suggest that gender accounts for some individual differences in mathematics achievement, other researchers report that other factors, such as age, type of task, or socialization experiences may better explain performance differences.

This paper begins by reviewing the available research on what is known about one noncognitive factor that may be associated with individual differences in mathematical achievement, namely gender. Possible reasons for identified gender differences in math are also reviewed. Additionally, tests used to assess mathematics performance are described. Finally, the results of a research study that investigated how gender relates to performance on standardized tests of mathematical skills of students in one school district in Illinois is described.

Overview of the Relationship Between Gender and Math Performance

Math continues to be viewed by some as a masculine domain. However, the research literature has found minimal differences in math test performance between boys and girls in elementary and secondary schools. When differences are found, they are more prevalent in higher grades (e.g., Tiedemann, 2002), among the highest achieving students (Garner & Engelhard, 1999) and on college-entry or other specific types of math problems or tests (Garner & Engelhard, 1999). The literature also shows a persistent advantage in math grades for female students from elementary school through the college years (Spelke, 2005). A review of this research follows.

<u>Age or Grade-Level of Student.</u> Many researchers have concluded that there are minimal differences between male and female students' math test performance in the elementary and secondary school years. For example, a group of researchers studying students in fifth and eighth grades in the United States did not find significant differences in mathematics achievement scores between boys and girls when measured with the California Achievement Test (Hall, Davis, Bolen, and Chia, 1999). Hall et al. noted that when found, gender differences often do not appear until late middle school and high school. A meta-analysis by Hyde, Fennema, and Lamon (1990) of 100 studies involving more than three million students reached similar conclusions. In elementary and middle school, girls outperformed boys on math tests by a small margin, but boys tended to outperform girls in the high school years, again by only a small margin. More recently,

Perie, Moran, and Lutkus (2005) reported findings from the Long-Term Trends in Mathematics National Assessment of Educational Progress (NAEP) study that indicated that the differences between boys and girls at all age levels were relatively small from 1973 to 2004. At ages 9 and 13, in 1973 girls outperformed boys on standardized achievement tests given in the schools by two points, but by 2004 boys were outperforming girls by three points. Among 17-year olds, boys consistently outperformed girls to a small extent, with the gap decreasing from eight points in 1973 to three points in 2004. Overall, results such as these have led to recent conclusions that girls and boys do not differ greatly in terms of math performance either during the elementary or secondary school years (Hyde, 2005; Perie et al., 2005; Spelke, 2005).

The greatest gender differences related to age are usually reported in favor of males on university entrance exams. Meta-analyses have found that male students tend to score higher than female students on the math portion of the Scholastic Aptitude Test (SAT-M) and ACT assessment test. Based on her review of meta-analytic studies, however, Hyde (2005) concluded that overall differences between males and females on the SAT-M are often small and the distributions of male and female scores are highly overlapping. Other researchers have suggested that the SAT-M under predicts the math performance of high school girls. When the SAT-M scores of male and female students were matched, for instance, women went on to earn higher grades in college mathematics classes (Spelke, 2005). In order words, women received higher grades in college math courses than their SAT scores would predict.

Greater Male Variability. The greatest gender differences on college entrance exams, such as the SAT-M, occur at the extremes of the test score distribution. At the high end of the test score distribution there is a greater tendency for male students to outperform female students (Hyde, 2005; Spelke, 2005). Spelke (2005) wrote that SAT-M scores suggest that either more boys than girls have extreme talent in mathematics or SAT-M scores overestimate the abilities of talented boys. There is also evidence that girls outperform boys at the lower end of the math achievement distribution. For instance, more boys than girls have been identified as having a learning disability in math (Spelke, 2005).

Grades Earned. An early report by Entwisle and Baker (1983) showed that there were no differences between the classroom grades of middle-class boys and girls in a first grade mathematics' class. While there were some differences for lower-class boys and girls, they were not consistent. These researchers also reported that boys often expected better grades, but did not consistently earn better grades. More recent research, however, suggests that girls and women earn higher grades in math courses in elementary and secondary schools, as well as at the college level (Spelke, 2005). For example, Kenney-Benson, Pomerantz, Ryan, and Patrick (2006) found that girls consistently receive higher classroom grades than boys do. Researchers speculate several causes for girls' higher grades in the classroom. Some have suggested that girls receive higher grades in the classroom because teachers have biased perceptions of the behaviors of boys and girls (Kenney-Benson et al., 2006), and these biased perceptions influence the grading of teachers.

There is also cross-cultural research to support higher grades for female students in mathematics. According to Muzzatti and Agnoli (2007), Italian women often outperform men in mathematics classes, both in high school and in college. Additionally, they are more likely to major in a mathematics field in college; however, they are less likely than men to find career opportunities within the mathematics field. Muzzatti and Agnoli suggested that societal factors, such as acceptance of high mathematics achievement in girls and women and the amount of career availability in mathematics for women, could be influencing both the way girls and women perform in the classroom as well as their not developing a career within the field of mathematics.

Types of Math Items. Garner and Engelhard (1999) found that when matched on ability, college men performed better in specific and different mathematic areas than women did, and vice versa. As a group, men displayed higher problem-solving ability and women had greater computational skills. Specifically, Garner and Engelhard found that men generally performed better in geometry and algebraic-arithmetic reasoning items, in addition to problem solving and other higher level cognitive tasks, including ratio, proportions or percents. Also, men performed better on items that contained tables, graphs or other figures. On the contrary, women were found to perform better on computational tasks, algebra, items using symbols and involving abstract mathematics, and tasks similar to those in textbooks. Spelke (2005) concluded that girls and women tend to excel on tests of arithmetic calculation and boys and men tend to excel on tests of mathematical word problems and memory for geometric configuration of the environment. Spelke also recognized that men and women tend to prefer different solution strategies. When a problem can be solved either by verbal computation or by spatial imagery, males are more likely to use the latter and they perform better on problems that benefit from this strategy than do women. Spelke also suggested that girls consistently perform better than boys on items that require one to determine if the data provided is sufficient to answer the problem. She reported that data-sufficiency items have been eliminated from the SAT-M and questioned whether the choice of items on college entrance exams and other standardized tests of achievement, such as the SAT, may be influencing the lower scores in math reported for female students in these tests.

Conclusions. Although a common stereotype may be that boys outperform girls on tasks involving arithmetic, as the research reviewed earlier indicates, this often is not the case. Female students tend to earn better grades in math classes throughout the academic years and male students tend to have an advantage on standardized achievement tests, especially college entrance exams, and to be over represented in samples of mathematically gifted and disabled students. However, findings related to gender and math performance may also be influenced by the age of the student and the type of solution strategy needed to solve the problems effectively.

Gender Similarities Hypothesis. Based on her comprehensive review of the meta-analytic research related to gender, Hyde (2005) proposed the gender similarities hypothesis. According to this hypothesis, "…males and females are similar on most, but not all, psychological variables" (p. 581). Hyde concluded that although an individual research study may have reported statistically significant differences for female and male participants, results from meta-analyses have indicated a pattern of gender similarities for math performance. Hyde also stressed that although gender differences are small and unstable for most traits and skills, the magnitude of gender differences can fluctuate with age. For instance, Hyde concluded that there are no gender differences in math computation in problem solving in elementary school ($d = .00$) or middle school ($d = -.02$), and that a small difference favoring males emerged in the high school ($d = +.29$) and college ($d = +.32$) years. Overall, Hyde concluded that we have overinflated claims of gender differences and that this may result in mathematically talented girls being overlooked by parents and teachers because these adults do not expect to find mathematical talent among girls. The next section describes these and other possible influences on mathematics' performance.

Possible Influences on Gender Differences in Math Performance Related to Gender

Gender differences in mathematic-related fields, such as physics and engineering, can be noted in the United States as well as other countries (Ma and Cartwright, 2003). Researchers have noted women are outnumbered by men in the areas of science and engineering occupations by five to one in North America. Other researchers have found that women hold only 10% of the jobs in physical sciences, engineering, and mathematics in the United States alone (Hall et al., 1999). Most college professors in the areas of mathematics, science, and engineering are men (Spelke, 2005). There are two possible reasons for this difference according to Hall et al. (1999): females have poor achievement in these areas or females have a negative attitude towards these areas of employment which result in them being less likely to pursue careers in these areas. Others have suggested a possible biological advantage for males to be better at math and science (Baron-Cohen, 2003). Ma and Cartwright (2003), however, have reported that these differences in the career choices of men and women have diminished over the last several years. Spelke (2005) further confirmed that in colleges in the United States, the academic pursuits of men and women are beginning to converge; approximately 47% of bachelor's degrees awarded in mathematics in 2000, for example, were awarded to women. Both Ma and Cartwright and Spelke attributed the differences in the world of work and in the perceptions of the mathematical potential for girls and women that do remain to social and cultural forces that create a negative attitude toward mathematics for women. It has been documented by several researchers that these negative attitudes may not be present until the later years of elementary school. The next sections describe what is known about how attitudes, anxiety, and biology may affect how female and male students perform mathematically.

Attitudes and Expectations of Students. Attitudes and expectations for success in mathematics have been found to vary by age and gender. In one study, for instance, a generally high, positive correlation between mathematics achievement and a positive attitude toward mathematics was reported for boys. However, when girls' responses were examined, these same correlations were low or negative (Schofield, 1982). Researchers have been unable to explain, however, whether positive attitudes are the cause of mathematics achievement, or if mathematics achievement causes the positive attitudes for boys (Newman, 1984).

When reviewing the literature, however, it is apparent that there is a trend for students in general to have more negative attitudes and higher anxiety towards mathematics as they progress through school (Fuchs et al., 2005). Ma and Cartwright (2003) discovered in their research that boys started seventh grade with a more positive attitude towards mathematics than girls, but both sexes' attitudes towards mathematics declined as they progressed to grade twelve. The rate of attitude decline was similar in both sexes, but the size of school had an effect on changes in attitude toward mathematics. Students in large schools tended to have a slower rate of decreased attitude towards mathematics, while those in smaller schools had a quicker rate of a decline in attitudes, suggesting that social factors such as diversity in teachers and course offerings could affect these differences. Ma and Cartwright pointed out that in larger schools students have more choice and may be able to select mathematics classes of more interest to them and this promoted more positive attitudes toward math.

Indications of gender differences in expectations for success in math were also found in an early study completed by Newman (1984). He found that girls in 10^{th} grade were likely to rate their self-perceptions of their own mathematics ability more poorly than boys. He also determined that a child's self-concept in second grade about his or her own mathematics ability did not predict mathematics achievement in later grades (fifth or tenth). However, the children who had high mathematics achievement in grade two also had high mathematics achievement in subsequent grades. As students aged, their self-perceptions of their mathematics ability were more likely to be positively correlated with their actual mathematics achievement. Researchers attribute the lack of predictive value of second graders' self-concept ratings to achievement in higher grades to the overestimation of their own abilities when younger. Older children's self-concept ratings were more veridical with their actual performance than were second graders'. Newman proposed that these findings also suggest that it is more valuable in lower grades to increase success in mathematics than to attempt to increase students' self-confidence regarding mathematics. However, more recent evidence has shown that students' predictions of mathematics performance do not differ between second, fourth or sixth graders, even when examined before and after students have taken mathematics tests (Barnett and Hixon, 2001).

Entwisle and Baker (1983) also studied children's expectations for their own performance in mathematics. While they generally found children's mathematics grades in the classroom to be similar for boys and girls, they reported differences in what the children held as expectations for their own mathematics performance. When comparing a lower-class school to a middle-class school, boys at both schools expected better grades in mathematics class than the girls did. However, considering that their grades were about equal, researchers attribute boys' higher expectations to their being overly optimistic or to girls being overly pessimistic.

Moreover, there is cross-cultural evidence that demonstrates a general decrease in self-confidence and value attributed to mathematics for both boys and girls as they advance through grades (Muzzatti and Agnoli, 2007). Muzzatti and Agnoli discovered that as Italian children of both genders progressed from elementary to middle school, their self-confidence in their mathematics ability seemed to decrease. These same children were asked which gender is better at mathematics; both the boys and the girls reported the boys to be better at mathematics than girls. In addition, both genders rated mathematics as being more difficult if they were in middle school than if they were in elementary school. This demonstrated a developmental variable that might affect how well boys and girls achieve in mathematics. Not only do both genders perceive that mathematics will become more difficult as they age, but they also believe that their chance of succeeding is less. Additionally, girls confidence in mathematics decreases while they continue to believe that boys' mathematics achievement is better then their own. Tiedemann (2002) suggested that girls are more likely to attribute failure to lack of ability than to lack of effort; they are also less likely to attribute success to ability. Both of these factors could explain lower mathematics expectations in girls.

Math Anxiety. Not only do negative attitudes toward mathematics increase with age and grade level, but so does anxiety about mathematical tasks (Kenney-Benson, Pomerantz, Ryan, & Patrick, 2006; Muzzatti & Agnoli, 2007) Math anxiety has been described as "...the negative emotional reaction that some people experience when placed in situations that require mathematical reasoning or problem solving" (Aschcraft, Krause, & Hopko, 2007, p. 329). This anxiety interferes with the manipulation of numbers and the solving of mathematical problems in a variety of life and academic situations. Math anxiety varies from mild states of apprehension or dislike to genuine fear or dread of mathematical activities. Highly math anxious students tend to earn lower grades in math classes and avoid math and math settings whenever they have the choice to do so. Even though there may be a conventional stereotype that girls and women are more likely to experience anxiety related to math, the limited empirical research on math anxiety and gender has not consistently supported this assumption. In their review of the available research, Aschcraft et al., (2007) noted that there is some evidence that girls and women are often overrepresented in high-anxious math groups and underrepresented in low-anxious groups. They further stated, however, that higher levels of math anxiety for females is typically not related to depressed performance on math tests or to greater mathematics avoidance. In fact, they reported that women now earn 50% of bachelor's degrees in science and engineering; 47% of natural science and mathematics bachelor's degrees; and 48% of graduate students in the natural, behavioral, and social sciences.

Home and Parental Influences. Where do attitudes and anxiety about math come from? One suggestion is that the home, specifically parents, may be one source. A great deal of evidence indicates that parents influence children's school performance and perceptions of competence, including mathematics achievement. Although they did not investigate how sex of parent or student might influence results, Hall, Davis, Bolen, and Chia (1999) studied middle school aged students and found that higher parental anxiety about mathematics was correlated with lower mathematics achievement scores for their children. Additionally, high mathematics achievement scores in students were positively correlated with parental exposure to higher level mathematics courses and with higher parental education levels in general. Parents' attitudes about mathematics and their beliefs in the utility of mathematics also related to higher mathematics achievement scores in students. Parents who had completed higher level mathematics courses tended to have more positive attitudes about mathematics, which seemed to influence their children's self-efficacy towards mathematics achievement. However, parental occupations and perceptions of their own math skills were not correlated with their children's actual mathematics achievement. When researchers separated black students from white students, these same trends did not appear for the black students as they had for the entire group; parental attitudes and mathematics courses completed by parents did not seem to be correlated with their children's self-efficacy or mathematics achievement.

The home environment itself has also been related to mathematics performance. For example, some research has determined that number of books in the home, availability of study aids, and parents' educational levels can predict student success in mathematics (Floyd, Evans and McGrew, 2003). Overall, it has been shown that the home environment can be critical for determining mathematics achievement in students.

Parents' stereotypes can also result in different attributions for boys' and girls' success in school. Studies have shown that parents have lower expectations for their daughters' math success than for their sons' (Hyde, 2005). Parents also tend to make different attributions for their boys' and girls' success and failures in math despite the fact that there is little to no differences in their actual math performance (Entwisle & Baker, 1983; Spelke, 2005). Parents of girls often report that their child is less talented in math than parents of boys. In addition, parents are more likely to attribute boys' success in math to talent and girls' success to effort. Parents usually attribute math failure to lack of effort for both boys and girls. However, mothers are more likely to attribute girls' failure to the task being too difficult. Overall, previous research indicates that parents appear less confident about their girls' than their boys' math abilities. As more women enter mathematics-related professions, it will be interesting to see how parental attributions related to girls' success in math might change over time.

<u>School and Teacher Influences.</u> One suggestion for why female students as a group do not do as well as male students on college entrance exams, such as the SAT-M, may be because girls are not encouraged and have not taken as many high-level math courses as their male counterparts. Recent research findings, however, indicate that girls and boys are taking higher level high school math courses with comparable frequency (Aschcraft et al., 2007; Hyde, 2005; Spelke, 2005).

Other researchers have suggested that teachers can influence how well students perform in the mathematics classroom. Several characteristics, including teacher training and the time a teacher spends on mathematics instruction, can predict mathematics achievement (Floyd et al., 2003). Also, the focus and content of curricular materials can be a factor (Floyd et al, 2003). Researchers have discovered that teaching behaviors related to lecture, explanation and instructional cues can have positive effects on mathematics achievement. Additionally, the more time teachers spend on lecture and explanation, the more positive students' attitudes become towards mathematics in general. Moreover, questions asked to large groups of students are more beneficial than questions asked to small groups of students, and waiting for student response or redirecting difficult questions also had a positive effect on mathematics achievement (Tomic, 1989). Teachers who have a high demand for correctness and provide little or no motivational support during class are at risk for promoting math avoidance in their students (Aschcraft et al., 2007). Because research in math has been less frequent than for reading, there is limited research describing how various teaching styles relate either to gender of teachers or students. There is some evidence that teachers have higher expectations for boys than for girls in math courses and they spend more time instructing, interacting with, and giving feedback to boys than to girls (Fennema, 1990). Female students may respond more positively to math instruction if it is taught in a cooperative manner, using an applied perspective and a hands-on approach rather than in the traditional competitive manner, using a book-learning approach (Fennema, 1990).

Biological Influences. Home and teacher influences are possible social factors related to how well someone performs in math. Another mentioned reason for the male-female gap on tests such as the SAT-M is that males have a biological advantage on tasks involving mathematics because of genetic, hormonal, and/or structural brain differences.

The cognitive skill that is most often mentioned that provides this biological advantage is spatial ability, which has been shown to be a significant predictor of math test performance (Spelke, 2005). It has been speculated that a male advantage in spatial ability is due to prenatal exposure to testosterone that influences brain lateralization and prenatal growth of the right hemisphere. Critics, however, argue that the research fails to support this interpretation. For instance, the link between math skills and spatial ability is based on correlational research. Additionally, gender differences favoring males are not found in all tests of visual-spatial ability (Spelke, 2005). The greatest male advantage is found in tests of mental rotation, but a female advantage is usually reported in tests of spatial visualization. In her comprehensive review of the literature on the development of math and science abilities, Spelke (2005) concluded that some individuals may be predisposed based on their biological endowment to greater achievements in math, but this variability in predisposition is not related to gender. Spelke stated that, "Research on the cognitive abilities of males and females, from birth to maturity, does not support the claim that men have greater intrinsic aptitude for mathematics and science." (p. 956). Instead, Spelke contends that mathematical and scientific reasoning develop from a set of biologically based cognitive capacities that males and females share, providing women and men with equal talent for mathematics and science.

Conclusions. Previous researchers have suggested that social factors, such as parents and teachers, and biological factors can influence students' attitudes, expectations and success in math. Which, if any, of these variables is more influential, and when and how they relate to a child's gender has yet to be agreed upon by researchers.

Standardized Mathematics Testing

Schools continue to use standardized tests to evaluate and make decisions about students, the curriculum, and quality of instruction. High stakes testing (standardized testing) has emerged with heightened importance in the field of education as a result of the No Child Left Behind Act (NCLB). Some researchers, however, have critiqued the value of high stakes testing in making decisions about children's education. For example, state standardized tests have been criticized for having poorly constructed questions and for using scales that do not accurately assess students' learning (Hursh, 2005). The overall reliability and validity of scores from standardized testing have been questioned (Hursh, 2005). Specifically, these scores have not been shown to assess adequate AYP for schools with diverse populations, such as racial and income diversities. For this reason, standardized scores may or may not be the best estimate of how much children in schools are learning. Despite these limitations, most states do use high stakes testing as a way to assess learning growth in students (Hursh, 2005). Therefore, research assessing trends in standardized testing is valuable. Two standardized tests that are used to assess mathematics achievement in the state of Illinois, the Illinois Standards Achievement Test and Prairie State Achievement Exam, are described next.

<u>Illinois Standards Achievement Test.</u> According to the Illinois State Board of Education (ISBE), the Illinois Standards Achievement Test (ISAT) is research based and broadly examines academic achievement in students. This is a group administered test given in the spring to children in public schools; it is used to measure progress in reading, mathematics, science and writing. Reading and mathematics are measured in grades three through eight, while science is assessed only in grades four and seven and writing is

assessed only in grades three, five, six, and eight. Questions for the ISAT were initially composed by and reviewed by teachers in Illinois (ISBE, 2003). Background characteristics of the teachers and how they were selected, however, are not specified. Questions were designed to match the Illinois Learning Standards for mathematics, which were also developed by Illinois teachers. There are eight mathematics standards that are tested using the ISAT. They include: Estimation/Number Sense/Computation, Algebraic Patterns/Variables, Algebraic Concepts/Identifying, Geometric Concepts, Geometric Relationships, Measurement, Data Organization/Analysis, and Probability. Some items on the ISAT are designed to cover multiple learning standards (Wick, no date). Initially, the accepted items relating to these standards were tested in actual classrooms. After revisions, pilot items accepted for yearly state-wide screening (Forness, 2007). There are five multiple-choice items, one short-response item, and two extended-response items that are piloted on each form of the test. Piloted items are screened for content, difficulty, precision, and fairness. Items that are too easy or too difficult based on pilot testing are omitted. Point-biserial correlations are used to determine the item's ability to distinguish between less proficient and more proficient students. Items with the highest point-biserial correlations are more likely to be chosen as actual ISAT problems. In addition, the ISAT mathematics section contains some norm referenced items from the *Stanford Achievement Test Series, Tenth Edition (SAT 10)* (Pearson Assessments), to compare students on a national level (Forness, 2003).

The mathematics section of the ISAT uses multiple formats including: multiple choice, short response, and extended response. All multiple choice items include four answer choices. Short response and extended response questions are scored by using a rubric. Specific rubrics are developed for each item. There are two short response questions which contribute to 5% of the student's mathematics score, and one extended response question which contributes to 10% of the student's mathematics score. For both, short response and extended response, the students' work and answer are scored. The scoring rubric for the short response questions is used to rate the response from 0-2 (0 = completely incorrect; 1 = partially correct; 2 = entirely correct answer and work). The scoring rubric for extended response items is split into three categories: mathematical knowledge, strategic knowledge, and explanation. Each of these categories has a scale from 0-4. Mathematical knowledge rates the students' ability to use mathematical concepts, while the strategic knowledge section is used to rate the students' ability to identify useful information within the problem. Finally, the explanation section of the rubric is used to score the students' ability to provide a justification for his or her work (Forness, 2007).

On the ISAT, there are three 45-minute sessions, and calculator use is allowed in grades four through eight. The first session contains 40 multiple-choice questions (30 are from an abbreviated form of the SAT 10). The second session contains 30 multiple-choice items and three short-response items. The final session contains two extended response items. Additionally, reference sheets are provided for grades seven and eight. These reference sheets include formulas for finding the area of plane figures and the volume of solid figures. Other formulas are included on the reference sheets, such as the Pythagorean Theorem.

Prior to 2006, scores were reported on a scale ranging from 120 to 200. Previously, "each scale [was] defined by letting 160 represent the average proficiency of the first-year test population. Every unit on the scale [represented] 1/15 of the standard deviation of proficiency scores for the first-year population" (p. 23). Therefore, the first-year mean was 160, while the standard deviation of the scaled scores was 15 for each grade level. "Results in subsequent years are equated to the base-year scale" (p. 23). This equating process made longitudinal comparisons possible. (ISBE, 2003) Starting with the 2006 administration of the ISAT, actual scores for the ISAT were reported on a standard score scale ranging from 120 to 411, regardless of the characteristics of the raw score distribution. The scale was changed to a continuous scale across grades three through eight to show the performance of students in all grades on the same scale. According to ISBE, "vertical scaling allows schools to better monitor and quantify students' progress across grades" (p. 6). In addition, the numerical values of cut scores that represent proficiency levels were also changed. Statistical bridge studies conducted in 2005 were used to set these new levels which reflect the old performance category scales, and according to ISBE, "require as much knowledge as (and no more than) the old ISAT scale did" (p. 6). (ISBE, 2006).

Students' scaled score reports for the mathematics section on the ISAT are categorized into one of the following: academic warning, below standards, meets standards, or exceeds standards. These categories are defined for each grade level on the Illinois State Board of Education website; descriptions of each category for every grade level include more information than can be concisely summarized in this paper. The Illinois State Board of Education (ISBE) gathered committees of "curriculum experts" to outline descriptions of the categories for each grade level. These individuals were educators from within the state of Illinois. To determine numerical cut-off scores for the categories, these educators were asked to indicate the percentage of students that would answer an item correctly from each of three groups: "just above the Academic Warning/Below Standards boundary, those who were just above the Below Standards/Meets Standards boundary, and those who were just above the Meets Standards/Exceeds Standards boundary" (p. 21). The reliability of these performance decisions was examined using two threshold-loss coefficients, p, "the proportion of persons consistently classified on two parallel tests" (p. 22), and k (kappa), "which corrects p for the proportion of consistent classifications that would be expected by chance" (p. 22). The p values for the ISAT ranged from .90 to .98. The kappa coefficients ranged from 0.62 to 0.85. (ISBE, 2007).

Additional psychometric properties of the ISAT can be found in the technical manual on the Illinois State Board of Education's website. The 2007 technical manual reported that internal consistency reliability estimates for total scores reported on the mathematics subtest ranged from .92 to .94 for grades three through eight. Also, the technical manual cites the percentage of inter-rater agreement for the mathematics extended-response items. These percentages for exact agreement range from 89% to 97% (ISBE, 2007). An independent study of the technical characteristics of the ISAT found consistent internal reliability around .95 for the ISAT mathematics subtest across three years (Wick, no date). The technical manual also lists the reliability of student performance decisions based on test scores. This refers to students placed in categorical levels referred to as Academic Warning, Below Standards, Meets Standards, and Exceeds Standards. The consistency of performance decisions ranged from $p = .90$ to $p = .99$, with kappa coefficients ranging from .62 to .85 for grades three through eight in mathematics (ISBE, 2007).

Prairie State Achievement Exam. The Prairie State Achievement Exam (PSAE) is given to students in eleventh grade. Similar to the ISAT, the PSAE also measures achievement in reading, mathematics, science, and writing. The PSAE is different from the ISAT in that the PSAE uses questions from the ACT exam developed by Pearson Assessments. The ACT Test battery has four multiple-choice sections: English, Mathematics, Reading, and Science. The ACT is typically taken by students seeking college admissions, loans, scholarships, and other awards. The mathematics portion of the ACT contains 60 items and has a 60 minute time limit. The development of the ACT was based on input from college faculty members from across the nation. The content is reviewed and periodically adapted to support national standards and curriculum.

The entire PSAE mathematics testing consists of a 60 minute session devoted to ACT questions and an additional 45 minute session comprised of the WorkKeys Applied Mathematics Test which was also developed by ACT. The WorkKeys assessment was designed to evaluate applicants for employment in American businesses. The development of WorkKeys is based on five phases: skill definition, test specifications development, prototyping, pretesting, and construction of operational forms. The WorkKeys mathematics section contains items ranging from five different skill levels, with Level 7 requiring complex mathematics skills, such as fractions and proportions, and Level 3 requiring basic mathematical skills, such as addition and subtraction. All items on the PSAE are multiple choice questions, and students may use calculators.

The PSAE technical manual can also be found on the Illinois State Board of Education's website. This manual suggests that the average internal reliability for the PSAE across several years is 0.91, demonstrating a high degree of stability over time. Similar to the ISAT, the PSAE cutoff scores are presented for the following categories: Academic Warning, Below Standards, Meets Standards, Exceeds Standards (ISBE, 2007). Finally, a longitudinal study determined that the overall correlation between students' test scores from 2003 on the ISAT and students' test scores from 2006 on the PSAE was 0.84 for mathematics. Although many schools present ISAT data broken down by gender, no statistical analyses could be found.

Score reporting for the PSAE is similar to that of the ISAT. Raw scores from the ACT Mathematics Test and the WorkKeys assessment are separately scaled into Z-scores using the means and standard deviations of the equated raw scores from a previous scaling group. The two Z-scores are then equally combined with the weight of 0.5 for each component to form PSAE raw scores. Finally, these scores (referred to as the PSAE raw scores) are transformed into scaled scores depending on the distribution of the raw scores. Similar to scoring on the previous version of the ISAT, the scaled scores on the PSAE can range from 120 to 200. The target mean and standard deviations for the PSAE scaled scores are 160 and 15, respectively. Some scaled scores are adjusted to meet these targets. For example, raw scores in the lower ends of the distribution are truncated to the scaled score of 120. The actual means and standard deviations from year to year vary due to rounding, truncations, and some score adjustments, but the actual means and standard deviations are generally very close to the targets. (ISBE, 2007)

Conclusions. Both the ISAT and PSAE provide information about how students in Illinois are progressing in math. The ISAT assesses students in third through eighth grade, while the PSAE tests the math skills of high school students in eleventh grade. School administrators use the results from both tests to determine if their students are meeting and/or maintaining adequate yearly progress (AYP). No statistical information is available, however, about whether or not there are gender differences in performance for either test.

The Present Study

Because of the need for schools to demonstrate students have achieved or maintained satisfactory progress, schools in Illinois administer the ISAT and PSAE regularly, despite the limited statistical and technical information available for both of these tests. Feedback from these tests is also used to evaluate the effectiveness of teaching and the curriculum. Teaching strategies and curriculum materials, however, are not the only factors that can influence students' mathematical performance. In some cases, gender may be related to performance on standardized measures of math performance. As reported earlier in this paper, students' attitudes and expectations about math performance, as well as parent and teacher influences may be related to a student's gender. The present study examined how gender relates to students' performance on standardized achievement tests regularly administered to students in Illinois. Based on previous research findings the following predictions were offered:

1. Because the ISAT is administered to students in grades three through eight, based on previous research findings (Hall, et al, 1999; Hyde, 2005; Hyde, et al., 1990;

Perie et al., 2005) no significant gender differences in mathematical performance were expected.

2. Because previous research suggested that gender differences favoring males in math performance may emerge in adolescence and be more pronounced on college entrance exams (e.g., Hyde, 2005), such as the ACT which is the basis of the PSAE, significant differences in the performance profiles of high school boys and girls were expected on the PSAE.

3. Because of reports of more boys than girls having a learning disability in math (Spelke, 2005), more boys than girls were expected in the academic warning category on both the ISAT and PSAE.

4. Because the greatest gender differences on college entrance exams (such as the ACT) occur at the high end of the test score distribution (Hyde, 2005; Spelke, 2005), more boys than girls were expected in the "exceeds standards" category on the PSAE.

5. Because there is a trend for students to have more negative attitudes and anxiety related to math as they progress through school (Ma and Cartwright, 2003), the higher the grade level, the smaller the percentage of students who were expected to be in the "meet" or "exceeds standards" categories on both the ISAT and PSAE.

Method

Participants

Data for this study were collected for 1,654 students (807 girls and 847 boys) from one school district in a Midwestern state. The school district had approximately 3,200 total students and met annual yearly progress (AYP) requirements at all grade levels (K-12) in 2007. Students in this district were approximately 93% white, with other minorities making up 7% of the school's enrollment at the time of this study. The overall drop out rate for this district was 3.8%. The low income rate was 39.2%. According to the Illinois State District Report Cards, students included in the low income rate were those who come from a family receiving public aid, students who qualify for free or reduced-priced lunch, and students who are supported in a foster home or live in an institution for neglected or delinquent children. This district's students overwhelmingly spoke English, with only 0.4% of students eligible for transitional bilingual programs.

Tests

Illinois State Achievement Test (ISAT). The mathematics overall score on the ISAT from the 2007 school year was located for 1,430 students (694 girls and 736 boys) in grades three through eight in the district. Reported scores on the ISAT can range from 120 to 400, and are based on the number of questions answered correctly by the student. These scores are reported on a continuous scale for students in grades three through eight.

The performance definition assigned to each student based on ISAT mathematics' performance was also recorded. Possible performance definitions assigned to a student included academic warning, below standards, meets standards, or exceeds standards for that grade level.

Prairie State Achievement Exam (PSAE). Overall scores on the mathematics section of the PSAE, as well as the performance definition assigned, was located for 224 students (113 girls and 111 boys) in eleventh grade for the 2007 test year. Scores for the PSAE can range from 120 to 200 (mean = 160; standard deviation = 15) and performance definitions included academic warning, below standards, meets standards, or exceeds standards. The PSAE is routinely administered only to students in the eleventh grade.

Procedure

The school district gave the researcher access to the Illinois Interactive Report Card website that contains all student test data for the district. Test scores, performance definitions, grade level, and gender information were all exported from this website for data analysis. The information was analyzed using a series of t-tests for independent means and chi-square tests with SPSS statistical software.

Results

Table 1 presents a summary of the mean ISAT and PSAE scores for male and female students. As predicted, a t-test for independent means confirmed that the ISAT scores for females students (M = 245.34, SD = 31.41) were not significantly different from the ISAT scores for male students (M = 244.08, SD = 31.46), $t(1428) = .76$, $p = .45$ (two-tailed). Contrary to prediction, the PSAE scores for high school girls (M = 159.09, SD = 14.54) were not significantly different from the PSAE scores for high school boys (M = 156.26, SD = 17.06), $t(222) = 1.34$, $p = .09$.

Gender and Categories of Test Performance

In order to investigate how overall category of performance designation (exceeds standards, meets standards, below standards and academic warning) related to gender, the percentage of students in each category was combined for the ISAT and PSAE (see Table 2). A chi-square test for independence revealed that, overall, gender and category of test performance were not significantly related, $\chi^2(3, N=1654) = 6.03, p = .06$, Cramer's $V = .06$.

In order to investigate if there was any gender difference associated with the academic warning category when results from the PSAE and ISAT were combined, follow-up comparisons using Bonferroni's correction were conducted (see Table 2). Results of this analysis confirmed that, as predicted, there were more male than female students in the academic warning category on the PSAE and ISAT combined, $\chi^2(1, N=1654) = 3.75, p = .03$, Cramer's $V = .05$.

<u>PSAE and Categories of Test Performance</u>. As is evident in Table 3, a larger percentage of male than female students was found in the exceeds standards, below standards, and academic warning categories, and a larger percentage of female than male students was in the meets standards category of the PSAE. In order to investigate if gender was significantly related to categories of test performance for the students in the eleventh grade, a chi-square test for independence was conducted on the percentages of males and females who scored in the various categories of the PSAE. Results confirmed a statistically significant relationship between gender and category on the PSAE, $\chi^2(3, N=224) = 13.68, p = .003$, Cramer's $V = .25$. Follow-up combined comparisons using chi-square tests corrected for continuity, however, indicated that, contrary to prediction, there was not a significant difference in the percentage of male and female students in the exceeds standards category on the PSAE, $\chi^2(1, N=224) = 1.98, p = .08$, Cramer's $V = .09$.

An additional comparison was made to determine how success (exceeds or meets standards) versus lack of success (below standards and academic warning) on the PSAE related to gender. Figure 1 shows that there were 66.4% of females who scored in the combined exceeds or meets standards categories, while only 48.6% of males scored in the exceeds or meets standards categories. Further, only 33.6% of females scored in the combined below standards or academic warning categories, while 51.4% of males scored in these categories on the PSAE. Chi-square tests corrected for continuity confirmed that there were significantly more female than male students in the combined successful category of the PSAE, $\chi^2(1, N=224) = 7.20, p = .007$, Cramer's $V = .18$.

Grade Level and Categories of Test Performance

Because ISAT was administered to students in grades 3 through 8 and the PSAE was administered to high school students in the eleventh grade, grade level comparisons of the percentage of students who were successful and unsuccessful could be made (see Table 4 and Figure 2). As predicted, results confirmed that grade level and category on the PSAE and ISAT were significantly related, $\chi^2(18, N=1654) = 170.71, p = .000$, Cramer's $V = .19$. Using a Bonferroni correction, follow-up analysis confirmed that students in the lower grades were significantly more successful in meeting or exceeding standards on the ISAT than were high school students taking the PSAE, $\chi^2(1, N=1654) = 112.59, p = .000$, Cramer's $V = .26$.

Discussion

The goal of this research was to examine how gender relates to performance on standardized mathematical achievement tests regularly administered to students in Illinois. As predicted, third- through eighth-grade students' scores on the ISAT did not vary by gender. This finding is consistent with previous research (Hall et al, 1999; Hyde, 2005; Hyde, et al. 1990; Perie et al., 2005). Specifically, Hall et al. (1999) found that there were no significant differences in scores between boys and girls on the California Achievement Test. A larger study conducted by Perie et al. (2005) found that from 1973 to 2004 differences in mathematical performance between boys and girls at all age levels were relatively small. As a result of meta-analyses and her review of the research literature, Hyde (2005) proposed the gender similarities hypothesis which suggested small variance between males and females on most psychological variables, including mathematical skills. ISAT results from the present study support the gender similarity hypothesis.

While Hyde's gender similarities hypothesis suggested little variance between math achievement for boys and girls, she did note that a small difference favoring males is likely to emerge in the high school and college years (Hyde, 2005). Hyde (2005) reported that the greatest gender differences are generally found on university entrance exams, such as the SAT and the ACT. Because the PSAE yields an ACT score, it was predicted that there would be significant differences in the performance profiles of high school boys and girls in the present sample. This prediction, however, was not confirmed. A recent large-scale study also reported similar findings. Hyde, Lindberg, Linn, Ellis, & Williams (2008) investigated the relationship between gender and mathematics performance with a nation-wide sample. They gathered high stakes testing results from ten different states (California, Connecticut, Indiana, Kentucky, Minnesota, Missouri, New Jersey, New Mexico, West Virginia, and Wyoming) and predicted that there would be gender differences favoring male students' mathematical performance in upper grade levels. They found, however, that there were no statistically significant gender differences at any grade level and therefore, there was no evidence for gender differences in mathematical performance favoring males that emerged in the high school years. In fact, effect sizes for gender differences were consistently found to be less than 0.10 when examining the test results of over seven million students on state assessments of mathematical achievement. Taken together, the present test results and those of Hyde, et al., (2008) suggest no significant gender differences at any grade level in the mathematical test scores from recent high stakes testing.

Previous researchers have reported, however, that performance on specific types of math problems may vary by gender. For example, boys often perform better in geometry, problem solving, higher level cognitive tasks, and imagery tasks, including graphs or figures, while girls tend to perform better on computational tasks, algebra, items using symbols, and calculation (Spelke, 2005; Garner and Engelhard, 1999). Because there was no statistically significant difference in overall test scores based on gender for third through eighth and 11^{th} grade students in the present sample, items on the mathematical portion of the ISAT and PSAE appear to assess the math skills for both boys and girls equally. Future research is needed to confirm this possibility.

Performance Profiles

Although results from the present study, as well as those reported recently by Hyde, et al., (2008), found no significant differences based on gender in mathematics achievement scores associated with high stakes testing, some gender differences were evident when examining performance profiles. Hyde, et al., (2008) reported that male students tend to have greater variance in the score distribution, which is supported by previous literature (Hyde, 2005; Spelke, 2005). For instance, Hyde, et al., (2008) stated that the ratio of boys:girls scoring above the 95th percentile and the 99th percentile were 1.45:1 and 2.06:1, respectively. Although in the present sample, there was approximately 1.6 boys for every girl in the exceeds standards category for the PSAE, the difference in the percentage of boys and girls in this category was not statistically significant. Although they described the ratio favoring males at the high end of the test score distribution as small, Hyde, et al., (2008) noted that according to their data, we would expect to see 67% of men in science, technology, engineering, and mathematics (STEM) careers and 33% women in these careers. Unfortunately, Ph.D. programs still generally have only about 15% women in STEM oriented programs.

In the present sample there was also evidence for more boys than girls being unsuccessful in their mathematical progress. As predicted, significantly more male than female students were found in the academic warning category when data from the ISAT and the PSAE were combined. When PSAE scores were divided into those indicating success (i.e., met or exceeded performance standards) and those describing unsuccessful performance (below standards and academic warning categories), significant differences favoring female students were found. In fact, slightly over half of 11th-grade male students failed to meet academic standards for mathematics on the PSAE, compared to approximately one-third of female students. There may be several reasons for the greater number of boys than girls in the academic warning category of the PSAE and the ISAT. Spelke (2005) noted that there is evidence that girls outperform boys at the lower end of the math achievement distribution, which could be tied to research that has identified more boys than girls being identified with a learning disability in math.

According to Fennema (1990), school and teacher influences on math achievement may benefit girls. For example, female students may respond more positively to math instruction that uses a hands-on approach. It has also been reported that females earn higher grades in math courses than males (Spelke, 2005; Kenney-Benson et al. 2006). Future research is needed to clarify why boys may be more at risk than girls for not demonstrating adequate progress in mathematics.

Grade-Level Differences

There were significant differences found in the present study when looking at grade level comparisons. As predicted, students in lower grade levels (grade three through grade eight) who took the ISAT were more likely to score in the exceeds or meets categories than students in grade eleven who took the PSAE. Previous researchers have suggested several reasons for younger students performing at a higher level than high school and college-aged students. For example, Ma and Cartwright (2003) found that both genders tend to have a decline in attitude towards mathematics as they progress through twelfth grade. Muzzatti and Agnoli (2007) noted that students tend to perceive mathematics as more difficult in higher grade levels than in elementary school. Floyd et al. (2003) reported that the time a teacher spends on mathematics instruction can predict mathematics achievement. The school district that was analyzed in the present study reported to the state of Illinois that they spend 52 minutes per day teaching mathematics in third grade and 44 minutes per day teaching mathematics in eighth grade (ISBE, 2008). Figures for time spent teaching mathematics in eleventh grade were unavailable. If students in the elementary classrooms are getting more instruction time in mathematics than students in secondary grade levels, then this could be a reason why fewer secondary students are achieving academic success on standardized tests in Illinois. This and other possibilities should be investigated so that methods that promote mathematics success at all grade levels can be implemented consistently.

Limitations

There are several limitations of the present research. First, data from only one school district in one state were analyzed to determine the relationship between gender and performance on standardized tests in mathematics. The results may not be applicable to all students.

Additionally, the school district that was studied did not have demographics that were representative of the state as a whole. The school district had a population that was 93% white, with all other minorities making up only 7% of the school's enrollment. However, in the state of Illinois, 54.9% of enrolled students are white, 19.6% of students are black, 19.3% of students are Hispanic, 3.8% of students are Asian/Pacific Islander, 0.2% of students are Native American, and 2.2% of students are multiracial (ISBE). Additionally, the school district studied had only 0.4% of students who were eligible for transitional bilingual programs, while the state of Illinois has 7.2% of students who are eligible for transitional bilingual programs. The district did match the state on two important variables. The low income rate for the district was 39.2%, which was close to the 40.9% low income rate of the state of Illinois. Additionally, the state of Illinois has a 3.5% high school drop out rate. Similarly, the district analyzed has a 3.8% high school drop out rate (ISBE). In order to obtain a more representative look at how mathematics' test scores are affected by gender and grade level on the ISAT and the PSAE, it would be beneficial to look at the data from many more school districts in Illinois.

Another limitation of this study is that only data from one academic year were analyzed. The Illinois Interactive Report Card which was used to obtain data is a new program that school districts can use to manage ISAT and PSAE data. Therefore, the data that were available were limited to what has been obtained since the program originated. For the district that was analyzed, data could only be obtained from the 2006 and 2007 testing years. Over time, the data available from Illinois Interactive Report Card should also increase. Analyzing how students perform on the ISAT and the PSAE from year to year could be insightful for a particular school district or for the state as a whole. For instance, an administrator would be able to determine where weaknesses in curriculum are, and if they are related to gender or grade level. Administrators at the district that was analyzed in the present study may be curious to determine why students taking the ISAT were more likely to achieve academic success than students taking the PSAE. Longitudinal studies could investigate the success of changes in curriculum and teaching methods. As noted previously, the district sampled in the present study reported spending less instructional time on mathematics as grade levels increased and how this relates to possible changes in achievement test success could be investigated longitudinally by following the same sample of students over time.

Although data from the PSAE and ISAT were combined for some analyses in the present study, we do not know if these two tests are comparable. For instance, it could be that that the two tests are not comparable in difficulty level and this is another possible limitation, especially when making comparisons across grade levels. This is an important concern, because of lack of information about the validity of both tests.

Future Research

Gender and grade level differences are only two variables that might be related to achievement test results in math. Future research could investigate differences in performance on standardized assessments based on race, socioeconomic status, disability, geographic location, or curriculum materials. Insights gained from such investigations could be beneficial to professionals working in the school district because the results may pinpoint areas of concern or areas where remediation and additional resources are needed. Additionally, standardized state assessments are only one source of the data for many school districts. Other assessment measures, such as district assessments, informal assessments, other standardized assessments results, and other student outcome measures could be compared to ISAT and PSAE scores to help school districts make predictions about how students may score on the state standardized assessments.

In addition, it may be valuable to look at the other subject areas of the ISAT and PSAE. Comparing these areas with the previously noted variables can also provide informative insight. Trends in data can be valuable information that can help educators align curriculum to needs of the student population. Future research should focus on these other factors in order to create a more descriptive picture of student achievement.

Because the results of high stakes testing are being used across the country as predictors of educator and student success, future research could also include analyses of the skill sets that are being tested. For instance, Hyde et al. (2008) coded the mathematic portion on the tests from all states where actual test items were available, using a four-level framework: Level 1 (recall), Level 2 (skill/concept), Level 3 (strategic thinking), and Level 4 (extended thinking). For most states, the researchers found that no item on standardized tests fell at levels 3 or 4. Garner and Engelhard (1999) showed that women were found to have greater computational skills, and men displayed higher problem-solving ability. If state tests focus on recall, skills, and concepts, it might be predicted that girls would show higher achievement than boys. The PSAE may not include items that allow boys to display their abilities to solve problems with strategic thinking. Hyde et al. (2008) cautioned that the United States may be falling behind other countries because many state standardized tests are not assessing students on their abilities to use complex cognitive skills such as reasoning, planning, and making connections between ideas and across content areas. Because many teachers across the United States are 'teaching to the test', leaving higher-order questions off of these tests may entice educators to neglect these skills during instruction.

Conclusion

Despite criticisms about their inability to assess students' learning accurately (e.g., Hursh, 2005), high stakes tests are currently used in the United States to measure student and teacher progress. Data from the present study found no significant differences in the mathematical test scores for boys and girls on both the ISAT and PSAE. However, when performance profiles were analyzed, results from the present study suggest that boys and students in high school are at risk for not demonstrating satisfactory progress in math achievement. Future research is needed to confirm these findings and to identify possible causes and successful interventions.

References

Ashcraft, M. H., Krause, J. A., & Hopko, D. R. (2007). Is math anxiety a mathematical learning disability? In D. B. Berch & M. M. M. Mazzocco, *Why is Math so Hard for Some Children?* (pp. 329-348). Baltimore: Paul H. Brookes.

Barnett, J. E. and Hixon, J. E. (2001). Effects of grade level and subject on student test score predictions. *The Journal of Educational Research, 90*(3), 170-175.

Baron-Cohen, S. (2003). *The essential difference: The truth about the male and female brain.* New York: Basic Books.

Entwisle, D. R. and Baker, D. P. (1983). Gender and young children's expectations for performance in arithmetic. *Developmental Psychology, 19*(2), 200-209.

Fennema, E. (1990). Teachers' beliefs and gender differences in mathematics. In E. Fennema & G. C. Leder (Eds.), *Mathematics and gender* (pp. 169-187). New York: Teachers College Press.

Floyd, R. G., Evans, J. J., and McGrew, K. S. (2003). Relations between measures of Cattell-Horn-Carroll (CHC) cognitive abilities and mathematical achievement across the school-age years. *Psychology in the Schools, 40*(2), 155-171.

Forness, M. (2007). *2008 ISAT Mathematics Presentation.* Illinois State Board of Education. Retrieved February 26, 2008, from http://www.isbe.net/assessment/isat.htm.

Fuchs, L. S., Compton, D. L., Fuchs, D., Paulsen, K., Bryant, J. D., and Hamlett, C. L. (2005). The prevention, identification, and cognitive determinants of math difficulty. *Journal of Educational Psychology, 97*(3), 493-513.

Garner, M. and Engelhard, G. (1999). Gender differences in performance on multiple-choice and constructed response mathematics items. *Applied Measurement in Education, 12*(1), 29-51.

Hall, C. W., Davis, N. B., Bolen, L. M., and Chia, R. (1999). Gender and racial differences in mathematical performance. *The Journal of Social Psychology, 139*(6), 677-689.

Hursh, D. (2005). The growth of high-stakes testing in the USA: Accountability, markets, and the decline in educational equality. *British Educational Research Journal, 31*(5), 605-622.

Hyde, J. S. (2005). The gender similarities hypothesis. *Americal Psychologist, 60*(6), 581-592.

Hyde, J. S., Fennema, E., Lamon, S. J. (1990). Gender differences in mathematics performance: A meta-analysis. *Psychological Bulletin*, *107*, 139-155.

Hyde, J. S., Lindberg, S. M., Linn, M. C., Ellis, A. B., and Williams, C. C. (2008). Gender similarities characterize math performance. *SCIENCE, 321*(5888), 494-495.

Illinois State Board of Education. (2003). *The Illinois State Assessment 2003 Technical Manual.* Retrieved January 17, 2008, from http://www.isbe.net/assessment/pdfs/isat_tech_2003.pdf

Illinois State Board of Education. (2006). *The Illinois State Assessment 2006 Technical Manual.* Retrieved November 19, 2008, from http://www.isbe.net/assessment/pdfs/isat_tech_2003.pdf

Illinois State Board of Education. (2007). *The Illinois State Assessment 2007 Technical Manual.* Retrieved November 19, 2008, from http://www.isbe.net/assessment/pdfs/isat_tech_2003.pdf

Illinois State Board of Education. (2007). *Prairie State Achievement Examination: Technical Manual.* Retrieved January 17, 2008, from http://www.isbe.net/assessment/pdfs/2006_PSAE_tech_manual.pdf

Illinois State Board of Education. (n.d.). *Student Assessment.* Retrieved October 2, 2007, from http://www.isbe.net/assessment/default.htm

Kenney-Benson, G. A., Pomerantz, E. M., Ryan, A. M., and Patrick, H. (2006). Sex differences in math performance: The role of children's approach to schoolwork. *Developmental Psychology, 42*(1), 11-26.

Ma, X. and Cartwright, F. (2003). A longitudinal analysis of gender differences in affective outcomes in mathematics during middle and high school. *School Effectiveness and School Improvement, 14*(4), 413-439.

Muzzatti, B. and Agnoli, F. (2007). Gender and mathematics: Attitudes and stereotype threat susceptibility in Italian children. *Developmental Psychology, 43*(3), 747-759.

Newman, R. S. (1984). Children's achievement and self-evaluations in mathematics: A longitudinal study. *Journal of Educational Psychology, 76*(5), 857-873.

Perie, M., Moran, R., & Lutkus, A.D. (2005). *Three decades of student performance in reading and mathematics*. Washington, D.C.: U. S. Department of Education.

Schofield, H. L. (1982). Sex, grade level, and the relationship between mathematics attitude and achievement in children. *Journal of Educational Research, 75*(5), 280-284.

Spelke, E. S. (2005). Sex differences in intrinsic aptitude for mathematics and science: A critical review. *American Psychologist, 60*(9), 950-958.

Tiedemann, J. (2002). Teachers' gender stereotypes as determinants of teacher perceptions in elementary school mathematics. *Educational Studies in Mathematics, 50*, 49-62.

Tomic, W. (1989). Teaching behavior and student learning outcomes in Dutch mathematics classrooms. *Journal of Educational Research,82*(6) 339-347.

Wick, J.W. (n.d.). Independent assessment of the technical characteristics of the Illinois standards achievement test (ISAT). *Illinois State Board of Education*. Retrieved January 17, 2008 from http://www.isbe.net/assessment/pdfs/wick_report.pdf

Appendix

Table 1

Score Summary for Students Taking the ISAT & PSAE

Variable	N	Mean	Standard Deviation
ISAT Females	694	245.34	31.41
ISAT Males	736	244.08	31.46
PSAE Females	113	159.09	14.54
PSAE Males	111	156.26	17.06

Table 2

Percentage of Female and Male Students in Each Category of Test Performance for ISAT and PSAE Combined

Category	Girls	Boys
Exceeds Standards	20.0%	20.7%
Meets Standards	64.3%	60.3%
Below Standards	14.9%	17.0%
Academic Warning	0.9%	2.0%

Table 3

Percentage of Female and Male Eleventh-Grade Students in Each Category of the PSAE

Category	Girls	Boys
Exceeds Standards	10.6%	17.1%
Meets Standards	55.8%	31.5%
Below Standards	29.2%	42.3%
Academic Warning	4.4%	9.0%

Table 4

Grade-Level Comparisons of Percentages of Students Who Fell Within the Successful and Unsuccessful Categories of Test Performance of the ISAT and PSAE

Category	ISAT (grades 3rd – 8th)	PSAE (11th grade)
Meets Standards and Exceeds Standards Combined	86.5%	57.6%
Below Standards and Academic Warning Combined	13.5%	42.4%

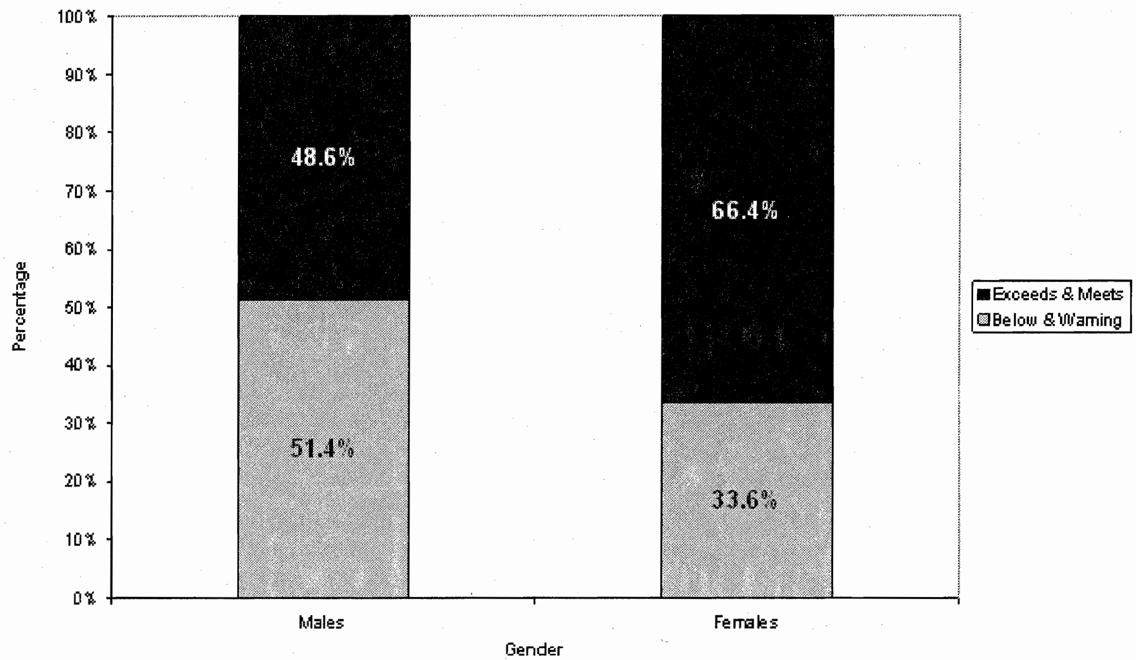

Figure 1: Gender comparisons of percentages of students who fell within the successful and unsuccessful categories of test performance on the PSAE

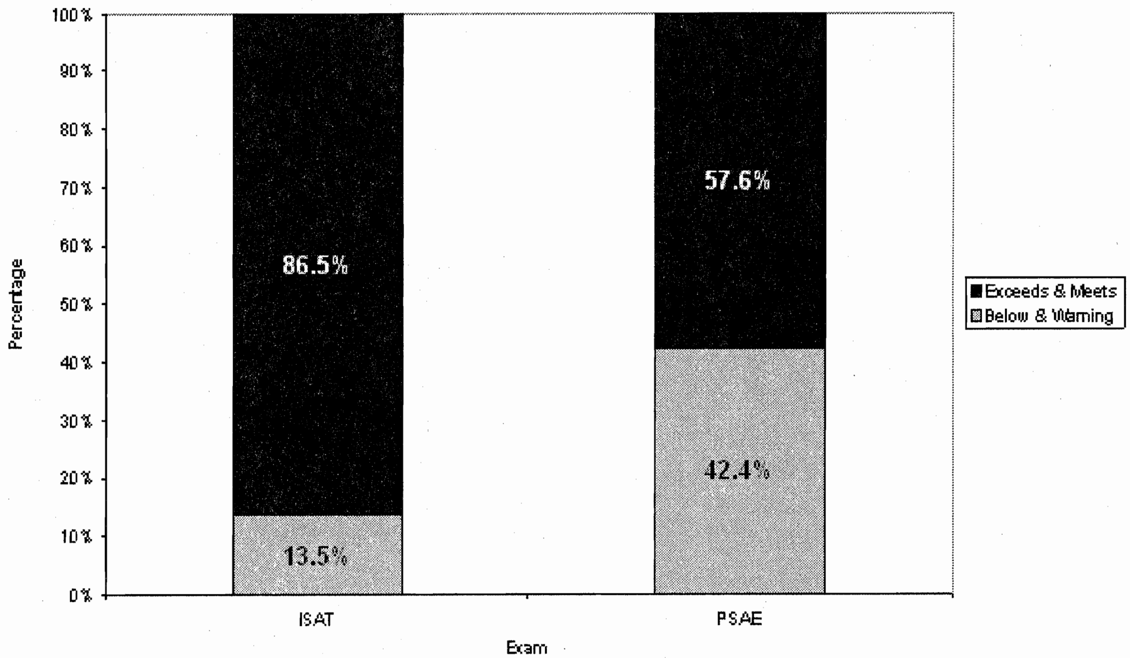

Figure 2: Grade-level comparisons of percentages of students who fell within the successful and unsuccessful categories of test performance of the ISAT and PSAE