8501114

Altenhof, Jody C.

INFLUENCE OF ITEM CHARACTERISTICS ON MALE AND FEMALE PERFORMANCE ON SAT-MATH

City University of New York PH.D. 1984

University Microfilms International 300 N. Zeeb Road, Ann Arbor, MI 48106

INFLUENCE OF ITEM CHARACTERISTICS

ON MALE AND FEMALE PERFORMANCE ON SAT-MATH

by

Jody C. Altenhof

A dissertation submitted to the Graduate Faculty in Psychology in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York.

1984

This manuscript has been read and accepted for the Graduate Faculty in Psychology in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy

7/16/84
date

Florence L. Denmark
Chairman of Examining Committee

7/30/84
date

Herbert D. Saltzstein
Executive Officer

Dr. Thomas F. Donlon

Dr. Samuel Messick

Supervisory Committee

The City University of New York

ABSTRACT

INFLUENCE OF ITEM CHARACTERISTICS ON MALE AND FEMALE PERFORMANCE ON SAT-MATH

by

Jody C. Altenhof

Adviser: Professor Florence Denmark

This study investigates the relative performance of males and females on certain item types on the SAT-Math. Samples of 1500 males and 1500 females divided evenly between three score groups (200-399; 400-599; 600-800) were selected from the January, 1978 and January 1979 administration of the SAT-Math.

The partial phi coefficient, which correlates item performance and sex, while controlling for total test score, was used to compare performance of males and females on algebra, geometry, arithmetic, miscellaneous math, regular math, quantitative comparison, male content and female content items. Males performed better on geometry, regular math and male content items than they performed on the rest of the test and the difference between each type of tiem and the remaining items were greater than the differences for females. Females performed better on algebra and miscellaneous math than they performed on the rest of the test and the difference between each type of item and the remaining items is greater than that of males.

The study also investigates the relationship between years of math and sex differences in math performance. A covariance analysis of sex differences in math performance controlling for years of math to math performance. Males consistently scored significantly better than females on total SAT-Math and all scales in the 600 - 800 range. While males generally scored better than females in the lower and middle ranges, the differences are usually not significant except in the lower range in 1978. Adjusted means for males and females indicate that equating for years of math may be a major factor in eliminating sex differences in math performance in the lower and middle ranges. Different male and female experiences with mathematics are discussed.

ACKNOWLEDGMENTS

I cannot pass up the opportunity to thank the many others who contributed to this work. I want first to thank Dr. Florence Denmark, my chairperson, and Dr. Samuel Messick, the two individuals who contributed most to my graduate education. Dr. Denmark's encouragement and guidance were invaluable in helping me to reach this goal. Dr. Messick's approach to psychological measurement provided an all too rare example of how to translate psychological theory into meaningful data.

Dr. Thomas Donlon consented to serve as a member of my committee although he was not affiliated with the City University. His substantive contributions and support went far beyond what was required. I want to thank all three of my committee members for their substantive contributions, cheerfulness and patience.

Many others contributed to this effort. Dr. Howard Ehrlichman and Dr. Sheila Pfafflin, my outside readers, offered provocative suggestions which contributed to the quality of the final study. The College Board graciously contributed the data without which this study could not have taken place. Dr. Larry Stricker, Mary Lynn Binckley and Ann Jungeblatt of Educational Testing Service provided technical analysis, technical support and administrative support respectively. Dr. John Martin wrote and processed the extensive computer programs and was a pleasure to work with. I am grateful to all of them. Finally, I would like to thank my mother, father and husband, Gary, who helped every step of the way.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER 1

INTRODUCTION

The issue of sex differences in mathematics performance has become a topical and controversial issues in psychology. Interest in this area has been prompted by the entry of vast numbers of women into the labor force and, more specifically, into the traditionally masculine areas of the labor force. Until recently, females constituted only a small percentage of students enrolled in higher mathematics courses and employed in math-related fields.

Sells (1976) poignantly illustrated the way in which mathematics serves as a filter which keeps women in a restricted range of occupations. After studying the entering class at Berkeley in 1973, Sells found that while 57% of freshman males had taken the four years of high school math necessary for eligibility to study calculus or intermediate statistics, only 8% of freshman females had completed the same high school curriculum. This information is noteworthy in light of the fact that calculus or intermediate statistics was required for all except five majors: education, guidance, music, humanities, and religion. As a result, 92% of the females, as opposed to 43% of the males, in that particular class at Berkeley were restricted to the abovementioned five majors unless they had taken the required high school math courses in high school or made them up in college.

In the most frequently cited review to date, Maccoby and Jacklin (1974) conclude that the following cognitive differences between the sexes are fairly well established: females have greater verbal ability than males and males excel in visual-spatial and mathematical ability. Maccoby and Jacklin state that these conclusions must be considered tentative since the number of studies in the area is small and much of the research is problematic. They add that there is no evidence that any performance differences which have been found represent innate cognitive differences and that there are many environmental factors which might account for these findings.

The reasons for performance differences in mathematics are still not clear. Most of the studies which show male superiority in mathematics leave a number of unanswered questions. Major studies in the area include the Project Talent study, the National Longitudinal Study of Mathematical Abilities (NLSMA), the National Assessment of Educational Progress (NAEP), the Fennema-Sherman Study and the Benbow-Stanley report of the Study of the Mathematically Precocious Youth.

Data for Project Talent was gathered in 1960 and reported by Flanagan in 1964. Project Talent findings showed negligible sex-related differences in Grade 9, but by Grade 12, males appeared to perform better. Project Talent made no attempt to control the number of math courses taken by their sample. It appears that Project Talent males had taken more math courses than females since they were enrolled in more math courses than females. It is likely, then, that the Project Talent study compared a population of males with more mathematical background with a population of females with less mathematical background.

The National Assessment of Education Progress data shows that males outperform females in mathematics at ages 17 and 26 through 35, but that at ages 9 and 13, sex differences in performance were minimal and sometimes in favor of females. Subjects for this study were selected by sophisticated random sampling techniques with no control for educational or mathematical background. Since males have traditionally taken more mathematics in high school than females, it is likely that the 17-year-old and 26- through 35-year-old males had taken more math than the females they were being compared with.

The National Longitudinal Study of Mathematics Abilities used sex as a control variable. Whenever significant interactions were found between sex and any other variable, separate analyses were done by sex. A summary statement shows "differences favoring girls were for variables at the comprehension level (the lowest cognitive level tested) and the differences favoring boys were for variables at the application and analysis level." Although the number of mathematics courses previously taken was controlled, the study did not report the size of the difference between the mean male and the mean female performance scores or discuss the educational significance of the results.

The Fennema-Sherman Study (1975) investigated a variety of levels of mathematical learning and a number of cognitive and affective variables hypothesized to be related to sex differences in mathematics performance. Their sample included male and female students in Grades 6 through 12 in four different school districts.

In Grades 6 through 8 (n = 1330) significant differences were found in favor of females in a low-level mathematical cognitive task in one school district. Significant differences were found in favor of males on a high-level cognitive task in another school district.

The Grade 9 through 12 sample (n = 1233) was made up of subjects whose mathematical background was carefully controlled. Significant differences were found in favor of males in two of the four school districts.

Benbow and Stanley (1980) studied mathematical aptitude prior to the onset of differential course-taking in 10,000 males and females by administering six different versions of the SAT-M to intellectually gifted junior high school students over an eight-year period. They found a large sex difference (.5 standard deviation) in favor of males in each of six separate studies over the eight- year period. Since subjects were students in the seventh and eighth grades, males and females had taken identical math courses. The researchers concluded that their results ruled out the hypothesis that male superiority in math was the result of males taking more math courses than females. Benbow and Stanley attribute their results to superior male mathematic ability, which may be related to greater male ability in spatial tasks. They then state that male spatial superiority may have both endogenous and exogenous causes. Since there is no evidence in their research which supports this conclusion, their interpretation of their results has generated a great deal of controversy.

SOCIALIZATION

Although all of these studies show sex differences in mathematics performance, there is little agreement on what causes these differences. While Benbow and Stanley attribute their results partially to "endogenous" or genetic causes, many researchers cite differing socialization factors as the reason for superior male performance in mathemtics. Fox (1977) broke down the most important socialization influences into five factors: 1) perception of career relevance of math; 2) influences of significant others; 3) perception of math as a male domain; 4) attitudes, self-confidence and values; 5) educational policies and practices.

Perception of Career Relevance of Mathematics

One reason that females may not achieve their full potential in math is that they fail to elect to take advanced courses in secondary school because they do not perceive them as useful to their future education and career plans (Hoover, 1972; Sherman and Fennema, 1976; Luchins, 1976; Hilton and Berglund, 1971; Fox, 1975c). Fennema and Sherman (1976) found significant sex differences in the perceived usefulness of math, but they found no sex differences in math achievement in the two high schools where males and females did not perceive differences in the usefulness of math.

Perceived usefulness of math appears to be different for males and females for three related reasons. First, females have been less oriented to career than males (Iglitzen, 1972, 1973; Ory and Helfrich, 1976). These differences have been found as early as kindergarten and first grade (Looft, 1971; Schlossberg and Goodman, 1972). Kaczala (1980) found that all students rated math as more useful for males than females. The final report of the National Institute of Education, Women and Math: Summary of Research Findings, supports the hypothesis that sex-role factors influencing attitudes toward math are most important in creating differences by sex in math performance. Second, females who express career interests are more likely than males to be interested in fields other than math and science (Rossi, 1965a, 1965b; Prediger, McClure, and Noeth, 1976). Even mathematically gifted seventh and eighth grade females differ drastically from males in their interest in careers in math and science (Fox, 1974b, 1975d; Fox and Denham, 1974). Finally, Fennema and Sherman (1976) have suggested that females are more unaware than males of the relevance of mathematics to many career areas other than purely scientific ones.

Influence of Significant Others

A number of studies show that interaction with "significant others" can result in a different quality and quantity of mathematical experience for males and females. These significant others include counselors, teachers, peers, and parents.

Counselors

Females are more likely than males to seek advice from counselors (Harway, Astin, Suhr, and Whiteley, 1976), but girls receive little encouragement from counselors to pursue math courses (Casserly, 1975; Harway, et. al., 1976; Christman, Vidulick, Drolle and Kirk, 1976; Haven, 1972; Luchins, 1976; Perucci, 1970; Schlossberg and Pietrofisa, 1974; Friedersdorf, 1970). Haven (1972) found that 42% of the girls who were interested in careers in math or science reported being discouraged rather then encouraged to take advanced math courses.

Teachers

Support or lack of support of teachers appears to have an effect on female achievement in math (Luchin, 1976; Ernest, 1976). Teachers, unfortunately, appear to have different expectations for females and males in math at the high school level. Teachers make more academic contacts with females in reading and with males in math (Leinhardt, Seewald and Engel, 1979). Becker (1981) found a higher quantity and quality of teacher contacts with male students in mathematics.

Peers

Adolescent females perceive real peer pressures <u>against</u> achievement in mathematics. Entwisle and Greenberger (1972) found that middle-class males were less supportive of achievement for females than were other females or males from the working class. The male to female ratio in the classroom also appears to influence both the number of math courses taken and math achievement.

Parents

Parental support and encouragement is crucial for females in deciding whether to choose mathematics courses in high school (Haven, 1972; Luchins, 1976; Poffenberger and Norton, 1956, 1959). Lack of parental support for females in mathematics manifested itself in a number of way. Parents are less likely to notice and encourage mathematical talent in females (Astin, 1974). Females receive fewer toys of an educational or scientific nature (Maccoby and Jacklin, 1974; Casserly, 1975). Parents also hold lower educational aspirations for daughters than sons (Levine, 1976; Casserly, 1975).

Aiken (1975) has proposed the related hypothesis that sex differences in mathematical ability are a partial consequence of same-sex modeling. Other studies have shown that the influence of fathers may be greater than mothers (Haven, 1972; Ernest, 1976; Helson, 1971; Block, 1973).

Perception of Mathematics as a Male Domain

Mathematics is generally believed to be a male domain. The two major hypotheses which have been offered to explain sex differences in math are the masculine identification hypothesis and the cultural reinforcement hypothesis (Aiken, 1970a, 1975, 1976; Astin, 1974).

The masculine identification hypothesis states that males and females who identify with their fathers or a generalized masculine sex role are better at math than those who have a feminine identification. This hypothesis appears to have originated from a report by Plank and Plank (1954) on case histories of women who were eminent in scientific and mathematical fields. Clinical interviews showed that these women identified with their fathers, but they did not score extremely high on the masculine identification scale of the California Personality Inventory. This contradictory evidence, as well as the fact that the sample is small and select, makes it difficult to draw any conclusions from this study in support of the masculine indentification hypothesis.

The cultural reinforcement hypothesis states that females are not reinforced for succeeding in mathematics because math is a male domain. Recent studies indicate that females do not perceive math as much as a male domain as they did before the women's movement, but that males still do (Sherman and Fennema, 1976; Farley, 1974).

Hawley (1971, 1972) found that women preparing for traditional female careers believed that significant men in their lives dichotomize attitudes and behavior into male/female categories, while those preparing for nontraditional careers did not believe this.

Television, textbooks and the rest of the media carry very clear messages about sex-role stereotyping and mathematics. Studies by Sternglanz and Serbia (1974) and Bergman (1974) indicate that females are underrepresented and generally appear in stereotyped roles.

Although a great deal of attention has been given to sex typing in children's readers, very little attention has been given to mathematics textbooks. Jay (1973) and Federbush (1974) found that mathematics textbooks portray males as active and females as passive. Rogers (1975) examined eight algebra texts and found that even algebra texts pictured men as alert and active while women are depicted as dull and insignificant.

Attitudes, Self-Confidence and Values

Sex differences occur in attitudes toward math at the college level (Aiken, 1972; Aiken and Dreger, 1961; Dreger and Aiken, 1957), but males and females do not differ in expressed liking for math compared to other school subjects in high school (Aiken, 1970b, 1976; Ernest, 1976; Fox, 1947b, 1947c; Stright, 1960). Whether sex differences in attitudes toward math are found in pre-college populations depends on the definition of attitude. Sex differences exist when attitude is defined as perceived usefulness of math, but not when it is defined as expressed liking for mathematics (Fox, 1975c; Hilton and Berglund, 1974). When attitude toward mathematics is defined as self-confidence in learning mathematics, there are striking differences between the sexes in both college and high school populations (Erlick and LeBold, 1975; Fennema and Sherman, 1976).

The relationship between attitudes toward mathematics and math achievement again depends on the definition of attitude. Perceived usefulness of math and self-confidence as a learner of math seem to be related to differing achievement in math. As far as expressed liking for math, only the extremes of feeling seem to be related to sex differences in math achievement.

Results from studies on the influence of fathers' and mothers' attitudes toward math on females' achievement in mathematics are unclear. It appears that fathers' expectations, but not fathers' attitudes or professions, are related to achievement and coursetaking for females. If fathers sex-type mathematics as a male domain, they are less likely to expect their daughters to achieve in math. The relationship between mothers' attitudes toward math and female math achievement is unclear, with some studies indicating strong relationships while other studies show no relationship.

Educational Policies and Practices

Programs for the gifted do seem to make a difference in female willingness to take further courses in mathematics. Casserly (1975) found that females who were enrolled early in special math programs reported that this early enrollment had influenced their decision to take advanced courses in high school. It is not clear whether all-female classes or schools promote greater achievement among females than mixed-sex groups, but it appears that advanced programs are most beneficial to females if the number of females does not become too small (Casserly, 1975; Fox, 1974a, 1976a).

CRITERION BIAS

Another possible reason for sex differences in mathematics performance is sex bias in the measure of mathematical achievement or the criterion used in the studies. Before undertaking a research program to understand differences in performance in mathematics between males and females, one must be sure that the criterion used to measure differences is, in itself, unbiased.

There are a number of different ways of determining sex bias in the criterion measure. Sex bias of different items on the criterion measure can be determined empirically or on the basis of inherent unfairness (Dwyer, 1981).

There are two methods to determine empirical bias. In the first method, items are administered to a sample of adequate size and are then examined for those items which show an item-sex interaction. Once these items have been identified, a professional subjective judgment is made as to whether the item-sex interaction occurred because of some property of the item which logically create bias against either males or females. A second empirical method to determine bias is to compare the means and variances of scores for females and males. If females score higher than males on one measure, but not on another, one must again decide subjectively if the differences are relevant to sex of the subject or an artifact of measurement.

<u>Gender-Content Bias</u>

There are a number of different kinds of bias which might occur in a test of mathematical achievement. The first is content bias. Content bias usually occurs as a by-product of test development. After the test is developed, items may or may not be reviewed for sex bias. Even if they are reviewed for sex bias and it comes to the choice of including an item with desirable item analysis characteristics or one which is sex-biased, preference is given to the item with desirable item analysis characteristics. While guidelines now exist for the elimination of sex "bias" in test materials, these guidelines are used with varying degrees of enthusiasm by test developers. Since they are generally not retroactive, there are still a significant number of tests in use which have not been reviewed for sex bias (Dwyer, 1981).

Dwyer (1981) has further noted that item content is related to item performance in a way that corresponds closely to traditional American notions of "masculine" and "feminine" interests. All other things being equal, males achieve higher scores than females on verbal items when the item content is primarily scientific, mechanical, business or practical affairs. Females typically obtain higher scores than males when the test or item content is drawn from the arts and humanities or is based on understanding human relationships. Murphy (1977) and King (1979) also noted that males scored higher on reading passages with practical and scientific content, while females scored higher on descriptive and imaginative passages. The actual research on the effects of item content is limited, but existing studies do show a relationship between sex-related content and differential difficulty by sex. Milton (1957, 1959) found that sex-related differences in test performance on mathematical problem-solving were related to sex-appropriateness of item content. A later study by Hoffman and Maier (1966) failed to replicate these findings, but Dwyer (1981) notes that Hoffman and Maier failed to make adequate distinctions between "masculine" and "feminine" content.

Coffman (1961) examined sixty verbal items in the 1954 SAT and identified nine items where differences between males and females were significant. Six of these item differences could be explained by sex-related content. The three items which favored females had "personal" or personality characteristics. One of the six items which favored males had mechanical content, while the other two had business content. Coffman also asked test specialists who constructed the 1954 SAT to predict which items would be easier for males and which would be easier for females. Item data was found to be in the predicted direction for fourteen of the seventeen items. Males performed better on six of the seven items judged to be easier for males and females performed better on eight of the ten items judged to be easier for females.

In one of the major studies in the field, Strassberg-Rosenberg and Donlon (1975) used delta plots to find items in the 1974 SAT with high item-sex interaction. They did this by taking delta values (an index of item difficulty) for males and females and cross-plotting then to get an elliptical pattern of points. The items departing most extremely from the concentration of points in the plots (outliers) are considered to have high item-sex interaction. An examination of the content of these items revealed that verbal items which were biased in favor of males contained traditional masculine content, while verbal items biased in favor of females contained traditional feminine content. In the math section subject content real world reference items showed an average bias in favor of males. Real world reference items include items which refer to real world objects such as cars, belts, etc., as opposed to human relations or ideas.

Mathematical Content Bias

A second type of bias in mathematical tests results from the choice of the type of mathematical content to be tested. Mathematical content may be divided into types of problems: algebra, geometry, arithmetic, etc., or may be divided according to level of problem-solving skill or abstraction required: conceptual problems vs. computational problems.

Test developers insure content validity by selecting items which cover the domain of mathematical content. There is, however, a certain amount of subjectivity in what constitutes the domain of mathematical content as well as the relative importance of various types of mathematical content. Another issue that could influence how items are selected from the domain of mathematical content is how different groups perform on these items.

If one group does better on conceptual problems than computational problems, the question arises of whether the test is biased toward that group if it contains more conceptual problems than computational problems. Likewise, if one group performs better on geometry than another group, it may be unfair to include more geometry than other types of problems or even, for that matter, to include geometry as a part of overall mathematical aptitude.

Donlon (1973) observed that females score relatively higher on tests of solving algebraic unknown than on tests of other areas of mathematics. Strassberg-Rosenberg and Donlon (1975) found that geometry items were "biased" on the average in favor of males, while algebra and miscellaneous math items were "biased" on the average in favor of females. "Bias" was defined as item performance relative to other items, not on absolute differences between the sexes. Items "biased" in favor of females, according to this definition, may be succeeded on more often by males. It is interesting to speculate why certain types of math content items would be "biased" toward males and certain types of math content items would be "biased" toward females.

Sex Differences in Spatial Performance

The most stable and reliable sex differences which have been reported are in geometry. Werdelin (1961) matched males and females on age, social class, reasoning, verbal and number abilities on a wide variety of geometric problems. He concluded that the differences in favor of males were attributable to inferior female skill in spatial visualization.

Some studies in spatial visualization show sex differences in performance (Hilton and Berglund, 1974; Reeves, 1973; Nevin, 1973; Simpson, 1974), while others do not (Jacobs, 1974; McClure, 1971; Merkel, 1974; Roberts, 1970). When differences do occur, however, they are always in favor of males.

Sweeney (1953) controlled for spatial relations in problem-solving tasks by matching groups on the Flags Test. The Flags Test is a test of spatial visualizations, devised by Thurstone, in which the respondent is presented with two faces of a flag and must indicate if they represent the same face of the flag or opposite faces of the flag. He found that spatial relations was a good control for eliminating sex differences in problem-solving as matching for mathematics. Among more recent studies, Fennema and Sherman (1977) found that males scored significantly higher in spatial visualization in two out of four high schools studied. Elmore and Vasu (1980) found that male graduate students received significantly higher mean scores than females on three out of five spatial visualization tasks. Sherman (1979) found that spatial visualization is a significant predictor of geometry performance for females and both sexes combined and math problem-solving for females. Spatial visualization is a significantly better predictor of geometry performance for females than males. Burnett (1979) found that when sex differences in spatial visualization were statistically controlled, there were no sex differences in math ability as measured by the SAT-M. On the other hand, Fennema and Sherman (1978) found that junior high males were not superior to junior high females in spatial ability, and Sherman (1980) found no sex-related differences in spatial visualization in Grades 8 or 11.

Biological Explanations for Different Male and Female Spatial Performance

There is also controversy about possible causes of sex differences in spatial visualization. Three different biological theories have been advanced: 1) females have lower spatial ability because it is transmitted by a recessive gene on the X-chromosome; 2) spatial ability is related to androgen level in both males and females; 3) males and females differ in timing of cerebral dominance which is related to spatial visualization.

Recessive Gene on the X-Chromosome

Stafford (1961) concluded that spatial visualization has a hereditary component which is transmitted by a recessive gene carried by the X-chromosome. According to genetic theory, the following pattern of family correlations on spatial visualization would have to exist in order for this to be true: a) father-son correlations should be approximately zero because males no not receive an X-chromosome from their fathers; b) father-daughter correlations and mother-daughter correlations will be greater than zero because females receive one X-chromosome from each parent; c) father-daughter correlations will be higher than mother-daughter correlations because females always receive father's only X-chromosome, but may receive either of their mother's X-chromosomes; d) mother-son correlations should be equal to father-daughter correlations because in both cases half of the X-chromosomal inheritance of the female member of the pair is identical to the X-chromosomal inheritance of the male member of the pair (Sherman, 1977).

Bock (1973) summarized results of studies by Stafford (1961), Hartlage (1970), Corah (1965), and Bock and Kolakowski (1973) on family patterns of correlations on spatial tests and found them to be consistent with that proposed by genetic theory. Bock states that this evidence for genetic sex linkage is persuasive, because correlations on verbal tests do not show the same pattern. Furthermore, Bock does not feel that performance differences between the sexes can be adequately explained by parental or environmental influences.

Results from other researchers have not confirmed the X-linked hypothesis. A large scale study in Hawaii of Americans and European ancestry and Americans of Japanese ancestry found that neither the scores of five different spatial tests nor the score of the spatial factor produces correlations consistent with the X-linked hypothesis (DeFries, Ashton, Johnson, Kuse, McClearn, Mi, Rashad, Vandeberg and Wilson, 1976). Williams (1975) did not find the predicted correlations on the Block Design test in 55 Canadian families. Bock (1967) found the distribution of spatial ability in a sample of 45 families inconsistent with the X-linked hypothesis. Loehlin, Sharan and Jacoby (1978) investigated families composed of a father, mother and two children to determine whether there is a sex-linked spatial gene. They found that parent-child and sibling correlations on a battery of eight cognitive tests of spatial, verbal, numerical and perceptual speed showed little, if any, evidence for the operation of the X-linked recessive gene postulated by Stafford. Analysis of the male/female score distributions provided results more favorable to the X-linkage hypothesis for the child generation. The authors point out that even if the X-linked recessive gene were shown to exist that it does not explain the whole male/female difference in performance on spatial tasks.

Recently, Benbow and Stanley (1981) resurrected the hypothesis of inherited male superiority in spatial visualization to explain why males scored so much higher than females in their study of mathematically talented youth. Benbow and Stanley offered no evidence to support their position other than conjecture.

Results from studies on the genetic theory of spatial visualization are inconsistent. Some studies show evidence which indicates genetic transmission of a spatial visualization gene on the X-linked chromosome while others do not. In the studies which indicate X-linked transmission of a spatial visualization gene, some researchers conclude that sex-linked transmission fully explains sex differences in performance in spatial visualization, while others conclude that sex-linked transmission only partially explains performance differences. In light of the contradictory evidence, it seems that other theories to explain sex differences in spatial visualization should be explored.

Androgen Levels

A number of studies have shown a relationship between androgen level and spatial performance. Broverman and Klaiber (1968) found that male performance on spatial tests was related to androgen levels, with higher androgen levels associated with higher spatial performance. Peterson (1973) found the same relationship for females at ages 16 and 19. Since Peterson's sample was small and her measures of androgyny have been questioned, the conclusions which can be drawn from her study are limited.

The most recent study on androgen level and spatial ability was conducted by Hier and Crowley (1982) who found that men with very low androgen levels at puberty performed poorly on spatial ability tests. They concluded that male sex hormones are necessary for the full development of spatial reasoning. Kagan (1982) criticized these conclusions in an accompanying editorial stating that it is possible that different levels of androgens are a factor in differences in spatial ability, but that androgen levels were not the sole factor or necessarily the most important.

Timing of Cerebral Dominance

Another hypothesis which has been advanced to explain male/female differences in spatial performance is that males and females differ in the timing of cerebral dominance or hemispheric specialization. Two opposing theories of hemispheric specialization have been proposed to explain sex differences in spatial skills. Buffery and Gray (1972) point out that spatial skills are stronger when they are bilaterally represented so that males' later and weaker hemispheric specialization facilitates their development of spatial skills. Levy (1972) makes the opposite prediction from the evidence that left-handed men (whose cerebral specialization is weak) obtain lower spatial scores. Levy hypothesizes that strong cerebral dominance enhances performance on spatial skills. She found that later maturers showed greater hemispheric specialization for speech perception than did early maturers of the same age and sex. She postulates that the maturational rate has certain physiological correlates that are probably present at an early age. Some of the correlates influence development of organization of brain functions and help to determine level of spatial ability. She suggests that late maturers show more hemispheric specialization of abilities than early maturers, and this difference in specialization mediates the effect of maturation rate on spatial ability.

Comprehension and Verbal Differences

Sex differences have also been found on two other types of item variables at the comprehension level and on SAT-M data sufficiency items. Sex differences favoring females were found in variables at the comprehension level in the National Longitudinal Study. These items were, for the most part, arithmetic items. Neither Donlon (1973) nor Strassberg-Rosenberg and Donlon (1975), however, found differences favoring females on SAT-M arithmetic items.

Both Donlon (1973) and Strassberg-Rosenberg and Donlon (1975) found that females did relatively better on data sufficiency items than regular math items. In data sufficiency items, two verbal or mathematical statements are given and a question is asked. The examinee is then asked to choose the correct answer from the following four statements:

a) Need only Statement A to answer question
b) Need only Statement B to answer question
c) Need both Statements A and B to answer question
d) Neither Statement A nor Statement B provides enough information to answer question.

There are two major characteristics which distinguish data sufficiency items from regular math items. Data sufficiency items do not require as much precision as regular math items, but do require more verbal skill. There is no empirical evidence which addresses the precision issue, and there is mixed evidence for female superiority on verbal items.

Until recently, female superiority on verbal tasks has been one of the most solidly established generalizations in the field of cognitive differences between males and females. While most studies of adolescents and adults show that females perform better than males on verbal tasks, the definition of verbal task and units of measurement are not comparable from one study to another. The usual advantage is .25 standard deviation, with a range from about .1 to .5 standard deviation. Only four of the twenty-four studies summarized in Maccoby and Jacklin (1974) showed that males performed better than females. Since all four of these studies took place in Great Britain, these results can probably be attributed to cultural differences.

Recent studies using the SAT-V and GRE-V have not supported the conclusion of female verbal superiority. When SAT-V scores from a random sample of high school seniors in 1966 were compared, there were no differences between males and females. A similar comparison of male and female SAT-V scores from May, 1967 to March, 1968 also showed no differences (Donlon, 1973).

Although the results of sex difference studies in verbal tasks are inconsistent and difficult to compare, all studies (except the four conducted in Great Britain) show that females perform either better or as well as males. Females' relative verbal skill could explain why females perform relatively better on data sufficiency than they perform on regular math items.

RATIONALE FOR STUDY

The preceding literature search suggests three specific item characteristics which seem to influence male and female performance in mathematics. These characteristics are mathematical content (algebra, geometry, etc.), item form (SAT regular math format and data sufficiency format) and gender content (male, female or neutral). The relative influence of these item characteristics and their relationship to each other has not been systematically studied. For instance, we do not know if math content has a stronger influence on differing male and female math performance than item form or gender content. Furthermore, we do not know if mathematical content influences male and female performance on certain item forms or items with certain gender content. Any systematic information on the influence of these item characteristics on male and female performance would be useful to test developers who wish to construct tests of mathematical ability which do not put a particular group at a disadvantage.

Since the SAT-M is so frequently used to measure mathematical aptitude and many of the preceding studies cited have used the SAT-M, it is the best instrument to use to further investigate these item characteristics. The present study uses the SAT-M to investigate the influence of these three item characteristics in a more systematic way than previous studies. The specific hypotheses based on the literature which were investigated are:

1) Males perform relatively better on male-content items than they perform on other items and that difference is greater than that of females

2) Males perform relatively better on geometry items than they perform on other items and this difference is greater than that of females;

3) Males perform relatively better on regular math items than they perform on other items and this difference is greater than that of females;

4) Females perform relatively better on algebra items than they perform on other items and this difference is greater than that of males;

5) Females perform relatively better on miscellaneous items than they perform on other items and this difference is greater than that of males;

6) Females perform relatively better on quantitative comparison items than they perform on other items and this difference is greater than that of males;

7) Females perform relatively better on female-content items than they perform on other items and this difference is greater than that of males;

8) When years of math is controlled, there are no differences in SAT-M scores of males and females;

9) Correlations between latest high school math grades and SAT-M are not significantly different for males and females.

These hypotheses are explained in greater detail in the following section.

CHAPTER II

Procedure

Test Form and Items

The 60 SAT-M items, total SAT-V score and SAT-V scale scores for antonyms, analogies, sentence completions and reading comprehension were taken from the January, 1978 (1Z ASA1) test administration and the January, 1979 (OZ BSA1) test administration. Relevant biographical data from the Student Descriptive Questionnaire, an inventory which accompanied these administrations, was also used. Data for this study was provided by the College Board.

These particular test administrations were chosen because test means and standard deviations in both years were similar (1978 SAT-M mean = 447, s = 117; 1979 SAT-M mean≈442, s = 113). Furthermore, it was assumed that the samples taking these tests were comparable since both tests were administered at the same time of year and were only one year apart.

Sample

Since the hypotheses to be tested concerned differences between males and females on certain subsets of scales of the SAT-M, an attempt was made to select the sample to eliminate differences between males and females on total SAT-M. A sample of 3,000 (1,500 males and 1,500 females) was selected from the 1978 administration and a second sample of 3,000 (1,500 males and 1,500 females) was selected from the 1979 administration. An attempt was made to equate males and females by selecting equal numbers of males and females from various score ranges.

Within each sample of 3,000, 1,000 candidates (500 male and 500 female) were selected with SAT-M scores of 200 - 399; 1,000 candidates (500 male and 500 female) were selected with SAT-M scores of 400 - 599; and 1,000 candidates (500 male and 500 female) were selected with SAT-M scores of 600 - 800. The sample was selected by taking every "n^{th}" candidate who took the SAT in each score group (200 - 399), (400 - 599), (600 - 800) in each test administration until 500 candidates had been selected. Because the total population in each subgroup varied, different "n's" had to be used to get 500 males and 500 females in each score group. The total number of candidates who took each test and the "n" used to obtain each sample of 500 candidates are presented below:

1978

SAT Math Score	Males on File	n	Females on File	n
200-399	6184	12	8395	16
400-599	12434	24	9398	18
600-800	3625	7	1175	2

1979

SAT Math Score	Males on File	n	Females on File	n
200-399	18681	37	24474	48
400-599	34341	68	25367	50
600-800	8631	17	2679	5

Although the sample was selected with the intention of eliminating score differences between males and females, this attempt was only partially successful. The mean for males in 1978 was 489.74 (s = 142.70) and for females was 480.33 (s = 139.39). Male mean in 1979 was 489.07 (s = 138.11) while female mean was 481.60 (s = 134.80). T-tests between males and females show p <.11 in 1978 and p <.13 in 1979. Male and female means for these samples are higher than the overall means cited in the first section of this chapter because proportionately more cases were selected from the higher score range as shown in the preceding tables. Means and standard deviations for males and females in each score range are shown in the following table:

1978

SAT Math Score	Male		Female		
	Mean	S.D.	Mean	S.D.	P Value
200 - 399	326.14	47.10	319.76	44.59	p <.03
400 - 599	488.44	54.75	479.60	54.11	p <.01
600 - 800	654.64	43.88	641.62	39.29	p <.001

1979

SAT Math Score	Male		Female		
	Mean	S.D.	Mean	S.D.	P Value
200 - 399	329.46	56.33	324.54	46.59	p <.08
400 - 599	490.48	56.33	484.80	54.19	p <.1
600 - 800	647.28	40.51	635.48	32.03	p <.001

The method of sample selection described here was the most reasonable and practical method to eliminate score differences between males and females. The only other alternative was to match males and females on SAT-M score. This was not done because it was too expensive and logistically very difficult.

Differences between males and females are significant in four of the six subgroups where the standard deviations are much lower than the total sample. The fact that this method of sample selection failed to eliminate sex differences on SAT-M is interesting in itself. It indicates that even when equal numbers of males and females are selected from various subgroups of the score range, it is not possible to eliminate sex differences in performance, but only to minimize them. The failure to select a sample which eliminates differences between males and females has implications for the way the study was carried out and how results were interpreted. These implications are discussed in the Analysis section of this chapter and Chapter III, Results.

Classification of Items

Each quantitative item was classified in three ways: mathematical content, item form, and gender content. Mathematical content classifications were algebra, geometry, arithmetic, and miscellaneous math. Item form classifications were regular math and quantitative comparison. Gender content classifications were male-content, female content, and neutral-content. The mathematical content and item form classifications are the standard classification codes used by the College Board, while the gender content classification was made by sorting items into typically male and typically female categories.

The mathematical content classification is based on arithmetic, algebra, and geometric concepts which are typically taught in elementary and junior high school years. Arithmetic includes the four basic operations of addition, subtraction, multiplication, and division; properties of odd and even integers; and percents and averages. Algebra includes linear equations, simple quadratic equations, factoring and exponents; it does not include the quadratic formula, fractional or negative exponents, or logarithms. Geometry includes the properties associated with parallel lines and measurement related concepts such as area, perimeter, volume, the Pythagorean theorem, and angle measurement in degrees. Knowledge of special triangles such as isosceles, equilateral and 30° - 60° - 90°, is also assumed. The basic geometric formulas required are printed in the descriptive booklet and are included for reference in the test booklet. Miscellaneous questions may contain properties of arithmetic, algebra, or geometry items, but are classified miscellaneous because they involve newly defined concepts or novel settings.

The regular math (RM) classification is the multiple choice format which is now familiar to most test takers. The quantitative comparison (QC) question is a novel format which uses built-in answer choices and which has replaced the data sufficiency item format. In the quantitative comparison item format, the test-taker is given two quantities and asked to choose one of the four standard answers shown below:

A. Quantity in Column A is greater

B. Quantity in Column B is greater

C. The two quantities are equal

D. The relationship cannot be determined by the information given.

In general, quantitative comparison questions require less time to answer, involve less reading, and require less computation than the usual multiple-choice question. The quantitative comparison item format replaced the data sufficiency item format which was used in the test from 1959 to 1974. Directions for the quantitative comparison question are considerably simpler than the data sufficiency questions. Individual quantitative comparison questions are generally less dependent on verbal skills than the Data Sufficiency questions. Statistical analysis has shown that replacement of data sufficiency questions with quantitative comparison questions has reduced the correlation between the verbal and mathematical score.

The gender content classification includes mostly items classified neutral because the College Board routinely screens SAT tests for items which are overtly sex-biased in content. Items were classified male content or female content by the author and the dissertation committee. Male content items include those items which contain male pronouns and content which was typically male such as sports or mechanical content. Female content items include those items which contain female pronouns or content which was typically female such as children going to school and prices in grocery shopping. Verbal items were classified according to standard College Board classifications: vocabulary, reading, antonyms, sentence completion, reading comprehension, and analogies. Only subscale scores were used. No item level data was reported for the verbal section.

The Student Descriptive Questionnaire (SDQ) is a biographical inventory which is administered along with the SAT. Certain item data was taken from the SDQ which seemed to be relevant to this study. These were:

Item 2 - Type of high school attended (public/private)

Item 3 - Type of high school program (college-prep, etc.)

Item 4 - Number of students in high school class

Item 5 - High school class rank

Item 7 - Number of mathematics courses student expects to take

Item 13- Latest grade in math

Item 18- Whether student will have completed advanced placement work in math before entering college

Item 39- Highest level of education completed by father

Item 40- Highest level of education completed by mother

Item 43- Parents' approximate income

Item 53- Self-assessment of mathematical ability

Analyses

Section 1

Influence of Three Item Characteristics on Males and Females Performance on SAT-M

A first stage of analysis was conducted to determine the influence of the three item characteristics of interest on 1978 SAT-M. Data were analyzed for differences between males and females on the following variables:

1) Mathematical Content - algebra, geometry, arithmetic, miscellaneous math

2) Gender Content - male content, female content, neutral content

3) Item Form - regular math, quantitative comparison

Previous research has indicated that the first two variables influence the differences in math performance between males and females. The third variable, item form, was investigated because previous research has shown relative differences in performance between males and females on regular math and data sufficiency items. The data sufficiency form is an item form previously used on the SAT - M which has since been replaced by an item form called quantitative comparison. While the data sufficiency form is basically different than the quantitative comparison form, the two forms are similar in that they are not standard multiple choice items and they require less precision than the regular math form. There is currently no research on differences between male and female performance on the quantitative comparison form.

In order to carry out comparisons of different types of items, subscales were formed in which each type of mathematical content was combined with each type of gender content and item form. Assuming that all subscales contain items, there should be 24 subscales - four math content x three gender content x two item form.[1]

The primary analysis for Hypotheses 1 - 7 was the partial phi coefficient. The partial phi coefficient is computed by taking each item's partial correlation with sex holding SAT-M constant. The partial phi coefficient represents the correlation of item and gender membership above and beyond the correlation of total SAT-M score and sex. This analysis was particularly useful in uncovering the items on which there were sex differences in performance over and above the differences between male and female on total SAT-M in this sample.

1 For the remainder of this paper, "subscale" will be defined as one of the possible 24 unique groups of items which can be formed by combining each of the four math content types with each of three gender content types and each of two item form types. Examples of subscales are: algebra/regular math/male content, or miscellaneous/regular math/female content. "Scale" will refer to all items of one of the math content types regardless of gender content and item form; all items of one of the gender content categories regardless of math content or item form and all items in one of the item form categories regardless of math content or gender content. There are nine scales: algebra, geometry, arithmetic, miscellaneous, male content, female content, neutral content, regular math, and quantitative comparison.

Partial Phi Coefficient

The partial phi coefficient was used by Stricker (1982) to examine verbal items on the GRE Aptitude Test for race and sex differences. Stricker found that the partial correlation index agreed with the item characteristic curve in particular items which functioned differentially for the sexes, but not in the particular items which functioned differentially for the races. Both of these indices consistently disagreed with the item difficulty index in both the proportion of items and the particular items identified as functioning differentially.

Operationally, the partial correlation index is computed by the following formula:

$$r_{iS \cdot T_c} = \frac{r_{iS} - r_{iT_c} r_{T_c S}}{\sqrt{1 - r^2_{iT_c}}\sqrt{1 - r^2_{T_c S}}}$$

where :

r_{is} is the correlation between the item responses (incorrect or omit = 0, correct = 1) and subgroup standing (male = 0, female = 1); r_{iT_c} is the correlation between the item response and the total score, adjusted for item overlap (the item is removed from the score in calculating the correlation) and corrected for attenuation in the score; $r_{T_c S}$ is the correlation between the total score and subgroup standing, corrected for attenuation by the former.

All these correlations are product-moment coefficients. In the common case where subgroup standing is a dichotomous variable, and the correlation between the item and subgroup standing is a phi coefficient, the significance of this partial correlation can be determined with the special test of significance for a phi coefficient, $X^2 = N\phi^2$, evaluated with one degree of freedom (Guilford, 1965). This test is an approximation because variables are not only partialed out of this phi coefficient, but also corrected for attenuation.

In the present study, partial phi coefficients were computed on every item of SAT-M. Items were then grouped into the scales discussed earlier and the number of partial phi coefficients which were significant in the male direction and in the female direction were computed for each scale.

In addition the following scale statistics were computed for each of the subscales:

1 - Means for males and females

2 - Standard deviations for males and females

3 - Coefficient alpha reliabilities for males and females

4 - Correlations for each of the scales with the test as a whole for males and females

The reliabilities of each scale indicate the scale's internal consistency. A scale must be internally consistent in order for it to be content valid. The limited research to date suggests that the internal consistency for the SAT-M may not be the same for males and females. Examination of scale reliabilities will indicate whether there are differences in internal consistency between males and females on the various scales of the SAT-M and provide information important for content validity.

Correlations of scales with the test as a whole provide information about convergent validity. By constructing various subscales of the SAT-M and examining performance differences between males and females, subscale intercorrelations can be examined to see if all the subscales measure the same construct for both males and females. If males score approximately the same on all the subscales, convergent validity will be demonstrated for males. If females score approximately the same on all subscales, convergent validity will be demonstrated for females. Discriminant validity will be discussed in Section 3.

Section 2

Student Descriptive Questionnaire

1- Effect of Years of Math on Male and Female Performance

One of the major variables of interest in the area of sex differences in mathematics is years of math training. Most studies have not controlled for years of math, and the few studies which have controlled for years of math show inconclusive results (Fennema-Sherman, 1975). When Benbow and Stanley controlled for years of math, they found that males performed better than females. Their sample was limited, however, to mathematically precocious youths.

The sample of the present study includes low and middle ability groups as well as high ability. The Student Descriptive Questionnaire gives information on years of math taken in high school from one to five years. Information from Item 7 of the Student

Descriptive Questionnaire will be used to examine whether there are differences in math performance between males and females when years of math is controlled over all three score ranges. In addition each of the scales in the three score ranges will be examined for differences between males and females when years of math is controlled. While there are typically problems with self-report data, available information on the Student Descriptive Questionnaire indicates that responses are generally accurate. The main reason for this accuracy is that the Student Descriptive Questionnaire is administered at the same time as the SAT and students seem to feel that their answers to the Student Descriptive Questionnaire will be checked. Random verification of Student Descriptive Questionnaire responses has indicated that the information is accurate. The major problem with this data is that there is no way to distinguish what kind of math was taken, only the number of years taken.

It is hypothesized that:

8) When years of math is controlled, there is no difference in SAT-M scores of males and females

2- Concurrent Validation with Grades in Math

In this part of the study males' and females' scores on SAT-M were correlated with self-report of high school grades in math. Self-reported high school gradesin math wereobtained fromitem 13 of the Student Descriptive Questionnaire. Item 13 asks students to indicate the "latest year-end or midyear grade received since the beginning of ninth grade". Students are also asked to further indicate if the grade was received in an advanced, accelerated or honors course. The major problem with this data is that latest high school math grade may not represent overall math ability as well as high school math grade average, but this information is not requested on the Student Descriptive Questionnaire.

Concurrent validation is obtained by correlating males' and females' scores on SAT-M with latest high school math grade. Both males and females with higher last high school math grade should have higher SAT-M scores. In addition, separate subscales were correlated with latest high school math grades to see if: 1) individual subscale scores correlate more highly or differentially with latest high school math grade than total SAT-M scores: 2) individual subscale scores correlate differently for males and females with latest high school math grade. If either total SAT-M test scores or individual scale scores correlate differently with latest high school math grade for males and females, further analyses will be necessary.
This would constitute evidence that SAT-M (or its subscales) and grades in high school did not measure the same construct.

It is hypothesized that:

9) Correlations between latest high school math grades and SAT-M are not significantly different for males and females

Section 3

Correlation of SAT-M with SAT-V

Scores from SAT-M were correlated with SAT-V in an exploratory analysis intended to provide information about whether scores on SAT-V and SAT-M relate differently for males and females. The following correlations were computed:

1) correlation of total SAT-M score with total SAT-V score for males and females;

2) correlation of scale scores on SAT-M with total SAT-V score separately for males and females;

3) correlations of six subtests of SAT-V (antonyms, analogies, sentence completion, reading comprehension, reading and vocabulary) with SAT-M score separately for males and females;

4) correlation of four subtests of SAT-V with scales of SAT-M separately for males and females.

These analyses provide information about discriminant validity of SAT-M with the SAT-V. Neither the total SAT-M score or the SAT-M scale scores should measure the same ability as the SAT-V or its scales for either males or females. Correlations of SAT-M scales with SAT-V and SAT-V scales were examined to insure that SAT-M scales correlate more highly among themselves than they do with SAT-V or SAT-V scales for both males and females.

Correlations of SAT-M with SAT-V may be potentially important in explaining any differences in performance between males and females which may arise on SAT-M or its scales. There are many hypotheses which might be advanced to explain how verbal ability might relate differently to SAT-M scores for males and females. A few of these hypotheses are listed here:

1) Verbal ability and mathematical ability represent two different cognitive processes and these processes correlate differently for males and females;

2) Scores on SAT-M are determined by verbal ability and one or more distinctly mathematical abilities. One sex may be higher on verbal ability and correspondingly higher on mathematical ability because they are using their verbal skills better on both the SAT-V and the SAT-M.

3) It may be that the same mathematical problem can be solved by different cognitive strategies. It may be, furthermore, that males and females tend to use different strategies when solving the same problem. Females, for instance, may employ predominantly verbal strategies where applicable to solve problems while males may choose from a number of different strategies.

Section 4

Prediction of Male and Female Mathematics Performance Based on Item Analysis of Math Type, Gender Content and Item Type

The purpose of the last stage of analysis was to predict male and female performance on hypothetical SAT-like mathematics tests of varying lengths from item analysis information obtained in the first stage of the study. These predictions are based on theory developed by Gulliksen in which he establishes relationships between certain scale statistics and total test statistics.

Means for males and females on scales of varying lengths were predicted by multiplying the average percent correct for each sex on items of that type from the original SAT-M by the projected number of items on the scale. Given, for instance, an average percent correct for females of .55 (10 correct out of 18) on the algebra items, one can predict the mean of a 40 item algebra test for females by multiplying 40 by .55.

Variances for scales of varying lengths were predicted by taking the point biserial correlation of each item with the scale score, multiplying each of these correlations by the standard deviation and summing. This sum was divided by K (the number of items in the actual scale) and multiplied by the number of items in the new scale.

Reliabilities for the new scales were obtained by using the Spearman-Brown formula:

$$r_{nn} = \frac{nr_{11}}{1 + (n - 1)r_{11}}$$

where r_{nn} is the reliability predicted for a test n times as long as the one for which data are available, r_{11} is the known reliability coefficient and n is the amount by which the test is increased.

The last analysis, then, is the prediction of the three item-based statistics discussed for some undetermined number of scales of varying lengths composed of items from each subscale of the three major variables of interest. Theoretically, there should be 24 such subscales since the first variable, math type, has four subscales; the second variable, gender content, has three subscales, and the third variable, item form, has two subscales. The actual data, however, did not contain items which could be classified into all 24 different subscales.

The intent of this stage of analysis was to arrive at information with which to predict male and female performance on hypothetical tests made up of any combination of three variables -- math type, gender content, and item form -- which add up to 60 items. This analysis provides information on how males and females would score if the SAT-M were made up of all one item type. Information from this analysis may be useful to test developers who wish to estimate how males and females would score if the number of items of a certain item type were increased or decreased.

In order to do this, a number of scales of varying lengths were simulated for each subscale. The first of these scales consists of 15 items -- the minimum number of items necessary for an acceptable reliability. Scales were then constructed of 25% (15 items), 50% (30 items), 75% (45 items), and 100% (60 items) of the 60 item test. The end point of this stage of analysis, then, is a 3 x 24 x 4 x 2 matrix [3 scale statistics (mean, variance, reliability), x 24 subscales x 4 lengths (15, 30, 45, and 60 items) x sex (male and female)].

Section 5

Replication

The final step in the analysis is to replicate findings from the 1978 SAT-M with a sample from the 1979 SAT-M. All the analyses described in the first four sections were performed on the 1979 sample as well as the 1978 sample. This replication insures that findings from the 1978 sample are not due to any bias either in sample selection or in that particular administration of the SAT-M.

CHAPTER III

Results

Subscales

Neither the 1978 nor the 1979 tests contained enough different types of items to make up the 24 possible subscales. Only 11 of the 24 subscales could be formed from the 1978 data, while 13 of the subscales could be formed with the 1979 data. The main reason for this was that very few items could be classified as male content or female content (1978 male content n = 3; female content n = 2; 1979 male content n = 6; female content n = 1). The 11 scales which were formed from the 1978 data and the 13 scales which were formed from the 1979 data are listed in Tables 1 and 2. Male content and female content items for 1978 and 1979 are listed in Tables 3 and 4. Miscellaneous items for 1978 and 1979 are listed in Tables 5 and 6.

Since the sample selection did not eliminate performance differences between males and females, it is important to examine individual scales for sex differences. Means and standard deviations were calculated for each scale and subscale and are shown in Tables 7 through 10.

Since male total SAT-M was higher than female total SAT-M, it was expected that male means would be higher than female means on most subscales. Male means were, in fact, higher than female means on all but three subscales [1978 regular math/sex neutral/miscellaneous math (male $\bar{x}$ = 1.65; female $\bar{x}$ = 1.76); 1978 quantitative comparison/sex neutral/miscellaneous math (male $\bar{x}$ = .41;

female $\bar{x}$ = .41); 1979 regular math/male content/arithmetic (male $\bar{x}$ = .56; female $\bar{x}$ = .57)]. Since there are only two items in the 1979 regular math/male content/arithmetic subscale, it is not possible to draw any meaningful conclusions about the results on this scale.

Organization of Results

Results are discussed in the following sequence. First, reliabilities for each scale are presented to show the internal consistency of each scale. Reliabilities can be found in Table 9. Next, hypotheses 1 - 7 are examined using the partial phi coefficient. Results for the partial phi coefficient are found in Tables 10 - 15.

The following other results are also discussed:

1) Covariance analysis which tests for differences between males and females on various scales while controlling for years of math (Hypothesis 8);
2) Concurrent validation of SAT-M with grades in math data from Student Descriptive Questionnaire (Hypothesis 9);
3) Correlations of each subscale with SAT-M (minus the subscale) for males and females in both years;
4) Correlations of SAT-M with SAT-V for both years;
5) Predicted means, variances, and reliabilities from Section 5 of Analysis Section.

Reliabilities

As noted above, reliabilities for both males and females for both years can be found in Table 9 [2]. Reliabilities were first computed for each scale using the number of items in the scale. They were then adjusted using the Spearman-Brown formula to reliabilities for 15 item scales. This was done so that scale reliabilities could be meaningfully compared.

Reliabilities are generally high and fall within the range of .7 to .86. Scales on which the reliability is below .7 include: 1978 quantitative comparison/sex neutral/arithmetic (female α = .69); 1979 regular math/male content/arithmetic (male α = .68; female = .68) and 1979 regular math/male content algebra (female α = .28).

Male and female reliabilities are usually quite similar. Scales where males and females differ by more than .03 include: 1978 quantitative comparison/sex neutral/arithmetic (male α =.76; female α = .69); 1978 regular math/male content/miscellaneous (male α =.71; female α =.28); 1979 regular math/male content/algebra (male α =.82; female α = .76). Since the last two scales originally contained only two items, it is hard to draw conclusions about reliability.

---

[2] Scales which have only one item are not discussed in this summary. These scales include: 1978 quantitative comparison/sex neutral/miscellaneous; 1978 regular math/male content geometry; 1979 regular math/male content/geometry; 1979 quantitative comparison/male content/algebra; and 1979 quantitative comparison/female content/arithmetic

Female reliability on 1978 regular math/male content/algebra is particularly low (α =.28). This scale contains two items -- item 18 and item 60, the last item on the test. Female mean on item 60 is .06. Only 817 females reached this item. Since so few females reached this item, reliability for females on this scale has little meaning.

Partial Phi Coefficients

Before discussing partial phi coefficient results for Hypotheses 1 - 7, it is necessary to specify the exact pattern of partial phi coefficients which would confirm the hypotheses. Hypothesis 1 is restated below in terms of partial phi coefficients:

> Hypothesis 1: Males perform better on items with male content as compared to all other items; moreover, whether or not females also perform relatively better on male-content items, the difference in performance between male-content items and all other items is greater for males than for females.
>
> Restatement of Hypothesis 1 in terms of expected partial phi coefficient results: A higher proportion of male-content items as compared to the proportion of all other items have partial phi coefficients which are significantly biased toward males; moreover, the difference between the proportion of male-content items with partial phi coefficients significantly biased in the direction of females and all other items with partial phi coefficients which are significantly biased in the direction of females is less than the difference between the proportion of male-content items with significant partial phi coefficients

which are biased toward males and all other items with significant partial phi coefficients which are biased in the direction of males.

Hypotheses 2 - 7 are similar to Hypothesis 1 in that each hypothesis states that one gender will perform better on a certain scale than all other items and that whether or not the other gender also performs relatively better on that scale, the difference in performance between the scale and all other items will be less for the second gender than for the gender first specified. The restatement of Hypothesis 1 in expected partial phi coefficient results can be used as a model for Hypotheses 2 - 7.

Hypothesis 1 : Males perform better on items with male content as compared to all other items; moreover, whether or not females also perform relatively better on male-content items, the difference in performance between male-content items and all other items is greater for males than for females.

This hypothesis is confirmed in both 1978 and 1979. Partial phi coefficients revealed that two of the three male content items (67%) are significantly biased toward males in 1978 and four of the six male content items (67%) are significantly biased toward males in 1979. 20% of the remaining items in 1978 and 19% in 1979 have significant partial phi coefficients which are biased toward males.

Moreover, none of the male content partial phi coefficients are significant in the female direction in 1978, and only one is significant in the female direction in 1979. 17% of the male content items have significant partial phi coefficients which indicate a bias toward females, and 19% of the remaining items have partial phi coefficients which were significant in the female direction in 1979.

Hypothesis 2: Males perform better on geometry items as compared to all other items; moreover, whether or not females also perform relatively better on geometry items, the difference in performance between geometry items and all other items is greater for males than for females.

This hypothesis is confirmed in 1978, but not in 1979. Partial phi coefficients indicated that three of the 18 geometry items (17%) are significantly biased toward males in 1978 and four of the 17 geometry items (24%) are significantly biased toward males in 1979. 22% of the remaining items in 1978 and 24% in 1979 have significant partial phi coefficients which indicate bias toward males.

There is one significant partial phi coefficient which indicates bias toward females in both 1978 and 1979. 5% of the geometry items and 21% of the remaining items have partial phi coefficients which were significant in the female direction. In 1979 6% of the geometry items and 23% of the remaining items are significant in the female direction.

Hypothesis 3: Males perform better on regular math items as compared to all other items; moreover, whether or not females also perform relatively better on regular math items, the difference in performance between regular math items and all other items is greater for males than for females.

This hypothesis is confirmed in both years. Partial phi coefficients indicate that ten of the regular math items (25%) are significantly biased toward males in 1978, and 11 of the regular math items are significantly biased toward males in 1979. 15% of the remaining items in both 1978 and 1979 have significant partial phi coefficients which indicate bias toward males.

Seven regular math items in 1978, and eight regular math items in 1979 have significant partial phi coefficients which indicate bias toward females. In the 1978 results, 17% of the regular math items and 15% of the remaining items have significant partial phi coefficients which indicate bias toward females, while in 1979, 20% of the regular math items and 15% of the remaining items have significant partial phi coefficients showing a bias toward females.

Hypothesis 4: Females perform better on algebra items as compared all other items; moreover, whether or not males also perform relatively better on algebra items, the difference in performance between algebra items and all other items is greater for females than for males.

In both 1978 and 1979, four of the algebra item partial phi coefficients are significant in the direction of females. 25% of the algebra items have partial phi coefficients which are significant in the direction of females in 1978 and 24% in 1979. In contrast, 13% of the non-algebra items have significant partial phi coefficients in the direction of females in 1978 and 16% of the non-algebra items in 1979.

Furthermore, four algebra items have partial phi coefficients which are significant in the direction of males in 1978 (25%) and three algebra items in 1979 (18%). 20% of the remaining items in 1979 have significant partial phi coefficients in the direction of males.

Hypothesis 5: Females perform better on miscellaneous items compared to all other items; moreover, whether or not males perform relatively better on miscellaneous items, the difference in performance between miscellaneous items and all other items is greater for females than for males.

In both 1978 and 1979, two partial phi coefficients on miscellaneous items are significant in the direction of females which represents 33% in 1978 and 25% in 1979. In contrast, 15% and 17% of the remaining items have partial phi coefficients which are significant in the direction of females in 1978 and 1979 respectively.

Furthermore, two partial phi coefficients on miscellaneous items are significant in the direction of males in both 1978 and 1979. In 1978, 33% of the miscellaneous items and 20% of the remaining items have partial phi coefficients which are significant in the direction of males. In 1979, 25% of miscellaneous items compared to 23% of the remaining items have significant partial phi coefficients in the direction of males.

Hypothesis 6: Females perform better on quantitative comparison items than they perform on other items; moreover, whether or not males perform relatively better on quantitative comparison items, the difference in performance between quantitative comparison items and all other items is greater for females than for males.

In both 1978 and 1979, three quantitative comparison items have significant partial phi coefficients in the direction of males and three in the direction of females. 1978 Partial phi coefficients reveal 15% of quantitative comparison items and 17% of remaining items are significant in the direction of females. In 1979, 15% of quantitative comparison items and 20% of remaining items have partial phi coefficients which are significant in the direction of females.

Furthermore, 15% of the quantitative comparison items and 25% of the remaining items in 1978 have partial phi coefficients which are significant in the direction of males. In 1979, 15% of the quantitative comparison items and 28% of the remaining items have significant partial phi coefficients in the direction of males.

Hypothesis 7: Females perform better on female content items than they perform on other items; moreover, whether or not males also perform relatively better on female content items, the difference in performance between female-content items and all other items is greater for females than for males.

Female-content items have no significant partial phi coefficients.

Relationship Between Partial Phi Coefficients and Item Difficulty

Item difficulties in Tables 10 and 11 show that, when SAT-M is held constant, items on which significant partial phi coefficients indicate bias toward males are more difficult than items on which significant partial phi coefficients indicate bias toward females in both years (1978 male-biased mean = 14.03; female-biased mean = 12.01; 1979 male-biased mean = 12.90; female-biased mean = 10.98).

Spearman rank-order correlation coefficients were computed to describe the relationship between the rank order of the significant partial phi coefficients and the rank of their item difficulty for males and females separately. There is very little relationship between rank of significant partial phi coefficients and rank of item difficulty for males in 1978 (r_s = .022) and 1979 (r_s = -.02) and females in 1978 (r_s = .05) and a very strong relationship between rank of significant partial phi coefficients and rank of item difficulty for females in 1979 (r_s = .82).

Summary of Partial Phi Coefficients

Results for Hypotheses 1 - 7 are summarized in Table 18. Hypotheses 1,3,4, and 5 are confirmed, and Hypotheses 6 and 7 have no support. Hypothesis 2 is confirmed in 1979 and partially confirmed in 1978. Although males do not perform better on geometry items compared to other items in 1978, males do perform better than females on geometry items. The difference between the proportion of geometry items and non-geometry items answered correctly is 8% for males and 16% for females.

Hypothesis 8: When years of math is controlled, there is no difference in SAT-M scores of males and females.

A covariance analysis using sex as the independent variable and years of math as the covariate shows no differences in performance between males and females. Mean scores of males and females for each year of math training (Table 19) indicate males and females in the sample have the same math training.

In order to more finely test this hypothesis, three separate covariance analyses were conducted to determine if, when controlling for years of math taken within each ability level, there are any differences in scores between males and females. Results from the covariance analyses with sex as the independent variable and SDQ7 (years of math taken in high school) as the covariate are reported in Table 20 for SAT-M and all the scales. A separate covariance analysis was done for each of these scales including the interaction (sex x SDQ7) which was found to be nonsignificant in all cases. Although the hypothesis was stated for the entire sample, the results suggest that when years of math is controlled, there are only differences between males and females in the 600 - 800 performance group.

While some results differ between 1978 and 1979, results from the two years are consistent in many respects. The most noteworthy finding is the striking difference between males and females in the 600 - 800 score range. With the three exceptions of miscellaneous math in 1978 (F = .18), female content in 1979 (F = .11) and arithmetic in 1979 (F = .53), all scales in both 1978 and 1979 are significant at the .05 level for the high SAT-M groups with males substantially outperforming females at the same level of high school math preparation.

Another important finding from these covariance analyses is that differences between the sexes in the middle ability range (400-599) are generally the smallest, followed by larger differences in the lower ability range (200-400) and the greatest differences in the highest ability range. The other important consistent finding is that with the exception of the 200-399 ability range in 1979, differences in geometry are consistently high across ability ranges in both years. In general, differences between males and females in the lower and middle ability ranges are much greater in 1978 than in 1979.

Adjusted means are reported in Table 21. Even when means are adjusted for years of math, males outperform females. The two exceptions in 1978 are miscellaneous math in the 200 - 399 score range and female content in the 400 - 599 score range.

In eight of the 24 cells in 1979, female adjusted means are equal to or higher than males. All of these differences are very small and none is significant. Female means for algebra, geometry, arithmetic and regular math in the 200 - 399 range and for algebra, miscellaneous, and female-content in the 400 - 599 score range are slightly higher for males than females. Male and female quantitative comparison means in the 400 - 599 range are equal.

Hypothesis 9: Correlations between latest high school math grades and SAT-M will not differ significantly for males and females.

This hypothesis is confirmed in both years. The correlation between latest grade in math and SAT-M score for both males and females in 1978 is .49. The correlation between latest grade in high school math and SAT-M for males in 1979 is .51, while the same correlation for females is .47. Correlations between latest high school math grade and total SAT-M and scales are shown in Table 22.

Subscale/Test Correlations

Correlations were computed separately for males and females between each subscale and total SAT-M minus the respective subscale. These correlations were corrected for attenuation in the SAT-M and are shown in Table 23. Most subscale/test correlations are approximately the same for males and females in both years.

Correlations which show the largest differences between males and females are: 1979 regular math/male content/miscellaneous (n = 2) male r = .66; female r = .58; 1978 regular math/male content/geometry (n =1) male r = .41; female r = .49; and 1979 regular math/male content/miscellaneous (n = 1) male r = .21; female r = .33. Since these scales contain so few items, it is quite likely that the male/female differences reported do not represent real differences.

Another finding worth commenting on is the difference in regular math/male content/ miscellaneous math from 1978 to 1979 (1978 male r = .60; female r = .58; 1979 male r = .21; female r = .33). Again, one must exercise caution in drawing any conclusion about inconsistency between the two years since regular math/male content/miscellaneous contains only one item in 1978 and two items in 1979.

Correlations Between SAT-M and SAT-V

Correlations between math and verbal scales are shown in Tables 24 - 27. The correlation between verbal and math is higher for females than males in 1978, and the same for males and females in 1979. In 1978 the correlation between SAT-M and SAT-V is .71 for males and .76 for females, while in 1979 the correlation between SAT-M and SAT-V is .74 for both males and females.

Correlations of geometry with verbal scales are lower than either algebra or arithmetic correlations with verbal scales for females in both years and males in 1979. Since algebra, geometry, and arithmetic have approximately the same number of items, these results may indicate that geometry measures an aptitude which is more distinct from verbal than either algebra or arithmetic.

Correlations between verbal and math scales do not show unusual differences between males and females. In both 1978 and 1979 SAT-M and math scales correlate higher with the reading scale than any of the other verbal scales and lower with the antonym scale than with any of the other verbal scales for both males and females. Regular math correlates more highly with the verbal scales than quantitative comparison for both males and females in both years. This can probably be attributed to the fact that the regular math scale is twice as long as the quantitative comparison scale and, therefore, more reliable. It is worth noting, however, that there is a larger difference between males and females in the correlation of regular math with the verbal scales than the correlation of quantitative comparison with the verbal scales.

Correlations for males and females for both years are somewhat lower on miscellaneous math with verbal and much lower on male and female-content with verbal. This is probably due to the lower number of items in these scales. The correlation of female content with verbal in 1979 is particularly low, but this scale contains only one item.

Predicted Means, Variances, and Reliabilities

The predicted means, variances, and reliabilties discussed in Section 5 of the Procedure section are presented in Tables 28 through 39 in Appendix. Reliabilities cannot be accurately predicted for all scales of one item and most scales of two items.

These tables provide information on the probable performance of males and females on each type of math item on scales of 15, 30, 45, and 60 items. Using information fromthese tables, the interested reader can predict probable male and female performance on 60 item SAT-M tests composed of various combinations of math item types.

CHAPTER IV

Discussion

Consistency With Literature

Findings from this study are basically consistent with Donlon (1973) and Strassberg-Rosenberg and Donlon (1975), the two major studies in this field. Because the three studies did not use the same methodology, it is difficult to make direct comparisons. Donlon (1973) looked at the difference between percent of males passing a certain item and percent of females passing the same item. Strassberg-Rosenberg and Donlon (1975) used delta-plots to examine item-sex interactions and noted items with strong item-sex interactions.

The primary analysis in the present study is the partial phi coefficient. The partial phi coefficient represents the correlation of sex with item response while controlling for total test score. Items could be biased either in the direction of males or females. The number of items which were significantly biased in each direction were counted for each scale.

Results from this study are most directly comparable to results from Strassberg-Rosenberg and Donlon (1975) since both analyses compare performance of males and females on subsets of items while controlling for total test score.

Findings Consistent with Literature

1- Males perform relatively better on verbal item content which is traditionally male (Dwyer, 1978; Strassberg-Rosenberg and

Donlon, 1974) and math item content with real world referents (Strassberg-Rosenberg and Donlon, 1975).

In the present study items were classified as male content if they contained a male pronoun or referred to "male" subject matter. Some subjects of male content items include: a mechanical belt, a swim meet, and a baseball team. All items which were classified as male content could also be classified as having a real world referent with the possible exception of the following item:

Jack begins reading at the top of page N and finishes at the bottom of page R. If the pages are numbered and read consecutively and if there are no blank pages, how many pages has he read?"

This item was classified male content because it contained a male pronoun. The subject content " reading a book" is neither male nor female, although it may actually have a more female tone since reading is a passive activity. This is the only item which contains a male pronoun, but no male subject content. Interestingly, this is the only "male content" item on which females scored better than males.

Because all the items classified as male content could also be classified as items with real world referents, results from this study can be compared with Donlon (1973) and with Strassberg-Rosenberg and Donlon (1975), which classified items in terms of real world referents. It should be pointed out, however, that while all the items classified as male-content in this study could also be classified real world referent items, the converse is not true. There are other items having real world referents which are not classified as male content. These items refer to subjects like " buying oranges", or "buying tickets to a play" or "studying either art or science".

Findings on male content items are consistent with Donlon (1973) and Strassberg-Rosenberg and Donlon (1975). A much higher percentage of partial phi coefficients are significant in the direction of males on male-content items than on non male-content items in both years. While there are no partial phi coefficients in 1978 on male content items which are significant in the direction of females, a higher percentage of partial phi coefficients on non-male content items than male content items are significant in the direction of females in 1979. In addition more partial phi coefficents are significant in the direction of males than females on male content items in both years.

2- Geometry items were "biased" on the average in favor of males (Strassberg-Rosenberg and Donlon, 1975).

This hypothesis is confirmed in 1979 by the partial phi coefficient criteria, but not in 1978. A higher percentage of geometry partial phi coefficients than non-geometry items are significant in the direction of males in 1979. Females, however, show a much different pattern than males on this criterion. A lower percentage of geometry partial phi-coefficients than non-geometry are significant in the direction of females in both years, but the percentage difference between significant geometry and non-geometry partial phi coefficients is much greater for females in both years than for males in 1978. Moreover, more geometry partial phi coefficients are significant in the direction of males than females in both years.

3- Regular math items were "biased" on the average in favor of males. (Strassberg-Rosenberg and Donlon, 1975).

A higher percentage of partial phi coefficients are significant on regular math than quantitative comparison items in the direction of both males and females, but the difference between the percentage of significant partial phi coefficients on regular math items minus the percentage of significant partial phi coefficients on quantitative comparison items is ten and thirteen percentage points for males compared to two and five percentage points for females.

4- Algebra items were "biased" on the averaged in favor of females (Strassberg-Rosenberg and Donlon, 1975).

This hypothesis is confirmed in both years. A higher proportion of significant partial phi coefficients on algebra items than non-algebra items are in the direction of females in both years. Furthermore, the difference between percentage of significant partial phi coefficients in the direction of males is less on algebra items than on non-algebra items in both years.

5- Miscellaneous math items were "biased" on the average in favor of females (Strassberg-Rosenberg and Donlon, 1975).

A higher percentage of partial phi coefficients on miscellaneous items than on remaining items are significant in the direction of both males and females. In both years, however, the difference between the percentage of partial phi coefficients on miscellaneous items minus the percentage of partial phi coefficients for non-miscellaneous items which are significant in the direction of females is greater than those in the direction of males. The number of partial phi coefficients on miscellaneous items which is significant in the direction of males is the same as the number significant in the direction of females in both years.

Findings Inconsistent With Literature

Findings from the present study are not directly comparable to earlier studies because the present study used quantitative comparison items and the findings in the literature refer to data sufficiency items. Data sufficiency and quantitative comparison item forms have some properties in common which distinguish both of these item forms from regular math. It is not clear, however, how quantitative comparison items and data sufficiency items are different.

6- Females did relatively better on data sufficiency items than regular math items (Donlon, 1973; Strassberg-Rosenberg and Donlon, 1975).

This hypothesis is not confirmed in either year using partial phi coefficients. Regular math items show a higher percentage of partial phi coefficients than quantitative comparison items in the direction of both sexes in both years. The actual number of significant partial phi coefficients in the direction of males is the same as in the direction of females in both years.

7- Verbal items relating to "things" (such as World of Practical Affairs and Science) are more difficult for females as compared to verbal items relating to "people" (such as Aesthetic-Philosophical and Human Relationships), which tend to be easier for females. (Strassberg-Rosenberg and Donlon, 1975).

Any conclusions about female performance on female content items must be very tentative for at least two reasons: First, there is less information on female performance on female-content math items than on male performance on male-content math items. Strassberg-Rosenberg and Donlon (1975) obtained mixed results in their subject-content analysis of math results and could not really draw a conclusion. The only evidence for superior female performance on female-content items comes from Strassberg-Rosenberg and Donlon's findings on verbal items. Secondly, the question of whether female subject content of math items has any influence on female performance on these items has not been adequately studied. The present study contains only three female-content items - two in 1978 and one in 1979. Part of the reason is that SAT-M has very few items which can be classified as female-content. The scarcity of female-content items can partially be explained by the effort to "sex neutralize" all items on the SAT-M. There were still, however, considerably more items in the two test administrations studied here and in the Strassberg-Rosenberg and Donlon study which were considered male-content than female-content. This may indicate some inconsistency between female content and mathematical problems in the minds of the test constructors.

This study shows no evidence to support a hypothesis that females perform better than males on items with female content. There were no significant partial phi coefficients on female content items in the direction of males or females in either year.

Major Findings

The strongest finding on math item types was a clear male superiority on male-content items. Other important findings were male superiority on regular math and geometry items and female superiority on miscellaneous and algebra items.

Male Content

Since there are still more male content items than female content items, further efforts should be made to "sex-neutralize" the content of the SAT-M. It is not clear at this point whether this means including more female-content items or making all items more sex neutral.

Geometry

This hypothesis was only partially confirmed in 1978 because males did not perform better on geometry items than other items. This finding is surprising and may partially be explained by the fact that geometry items were more difficult than other items.

On balance, however, the hypothesis was confirmed, raising an interesting issue. The present sample was selected so that there was minimal difference in male and female total SAT-M score. Furthermore, in both years there was very little difference in the correlation between grades in math and SAT-M score for males and females (1978 - r = .49 males, r = .49 females; 1979 - r = .51 males, r = .47 females). These correlations indicate concurrent validation for males' and females' overall math performance on a sample which was selected to minimize performance differences between males and females. With this in mind, the glaring inequality of male and female performance on geometry items suggests a source of bias against females. Geometry is, however, an important part of mathematics, and one which should be measured by the SAT-M. A better way to deal with geometry items might be to include them in a section which is separate from the rest of the SAT-M in much the same way as the writing sample was once separated from the SAT-V. In this way, the SAT-M would not contain a scale which was clearly biased against females, but geometry aptitude would still be measured.

Miscellaneous Items

While females did perform relatively better on miscellaneous items than other items, female performance on miscellaneous items is not unusually strong. The influence of item difficulty on female miscellaneous items has been noted earlier. Female performance on miscellaneous items in this study is probably also distorted by the fact that three of the fourteen miscellaneous items are male-content items while none are female-content items. The three male-content items are three of four items on which males scored highest relative to females. The male content of these items makes the findings on the miscellaneous item type difficult to interpret. The strong showing of males on other male-content items indicates that previous findings of relative female superiority on miscellaneous items might have been supported more strongly if the items in these particular tests had been more sex neutral.

Three of the four items most biased toward females were most like algebra and the remaining item was most like arithmetic. Otherwise, these items had no discernible properties in common.

Regular Math and Algebra Items

Other important findings are the strong showing of males on regular math items relative to quantitative comparison items and of females on algebra items relative to other items.

Female-Content Items

It is interesting to speculate as to why females did not seem to perform any better on female content items than on any other type of items. One reason that it is hard to draw any conclusion is that there are only three female content items in the two tests combined. Another reason is that all the female content items came at the end of the test -- the 53rd and 57th items in 1978, and the 51st in 1979. In two of the three cases more females did not reach the item than males, although on the third case, more males than females did not reach the item.

Why would male content items facilitate male performance while female content items do not facilitate female performance? Further research is necessary to answer this question, but the socialization literature raises some hypotheses. If math is considered a male domain, females may experience a form of "cognitive dissonance" which interferes with problem-solving abilities when they approach a math problem with a female agent. If females are indeed discouraged from taking mathematics by counselors and peers, and ignored, relative to males, by math teachers, it is reasonable to think that at some level, females would question the appearance of a female or female content in a math problem.

It is also interesting that males did not perform significantly better than females on female content items. Since males did perform significantly better on many of the scales, the data suggests that female content does not help females, but may hinder males.

Relationship of Years of Math to Sex Differences in Math Performance

The most striking finding in this study is that ability levels moderate the relationship of years of math to math performance. Most studies in the area of sex differences in mathematics have not controlled for years of math training. The major exception is the Benbow and Stanley study of seventh and eighth grade males and females with the same amount of formal training in mathematics. Benbow and Stanley found male performance was clearly superior to female performance. They concluded that their results ruled out the hypothesis that male superiority in math was the result of males taking more math.

The present study helps to explain Benbow and Stanley's results. Benbow and Stanley used only high ability males and females and found tremendous differences. The present study indicates that the greatest differences between males and females exist at the highest ability level when years of math are controlled.

This finding holds even when the differences between males and females in the present sample are considered. The actual differences between males and females in the high score range are highly significant in both years ($p<.0001$ in 1978 and 1979). When years of math is controlled, males and females show the greatest difference on algebra, geometry, and regular math in both years. The substantial difference in algebra is surprising. Sex differences in the low and middle score range, however, are not very great, and, as indicated in the in the table of adjusted means, are in the direction of females.

In 1978 males and females performed significantly different in the low score group ($p<.03$) and in the middle score group ($p<.01$). In the original sample, then, there was actually less difference between males and females in the middle score range than the lower score range. When years of math is controlled, the differences on all the scales (except miscellaneous) are greater in the lower score range than the middle score range. In both the lower and middle score range female miscellaneous math adjusted mean is higher than that of males.

In 1979 differences between males and females in the original sample are slightly less than in 1978 (200-399 $p<.08$; 400-599 $p<.1$). Without controlling for years of math, differences between males and females are slightly greater in the lower range than in the middle range in 1979. When years of math is controlled, males perform significantly better in geometry ($p<.03$) and much better in regular math ($p<.18$). The surprising finding in 1979 is that, when years of math is controlled, females seem to perform slightly better in geometry and regular math in the lower score range. The adjusted mean for males in geometry is 4.48 and for females is 4.49, while the adjusted means for males in regular math is 10.96 and for females is 11.07.

Results from this study, then, lead to a different conclusion than that drawn by Benbow and Stanley. Benbow and Stanley conclude that there are substantial sex differences in math performance when number of math courses is controlled. Results from this study indicate only that there are substantial sex differences in math performance for high ability groups when number of math courses is controlled. Even when sex differences in the original sample are taken into account, males perform substantially better than females with the same math training only in the high ability group.

Adjusted means are reported in Table 20. Even when means are adjusted for years of math, males outperform females in 1978 on all but two scales. The two exceptions in 1978 are miscellaneous math for the 200 - 399 range and female content for the 400 - 599 range.

Female adjusted means were equal to or higher than male adjusted means in eight of the 24 cells in 1979. All of these differences are very small and none is significant. Female means for algebra, geometry, arithmetic, and regular math in the 200-399 range and for algebra, miscellaneous math and female content in the 400-599 score range were slightly higher for males than females. Male and female quantitative comparison means in the 400-599 range were equal.

When interpreting these results, it is important to remember that a large number of students in the sample did not answer the SDQ item which asks how many years of math they took. 511 males and 412 females in 1978 and 501 males and 387 females in 1979 did not answer. Approximately one third of both sexes did not report years of math in both years. These subjects could not be included in the covariance analysis discussed in this section.

It is important to note one aspect of the sample who did answer SDQ7. In most cases the female mean for years of math is slightly higher than the male mean. Since it was possible to answer this question by responding "did not take Math", it cannot be assumed that those who did not answer SDQ7 did not take math. The data suggest that either more males with higher scores per each year of math did not report scores, or that females actually score slightly better than males with the same amount of math training. It is impossible to accurately interpret Table 20 and its implication for the data on years of math training without the data for those who did not answer this item. It would clearly be desirable in further research to have information on years of math taken for each person in the sample.

Taking into account the possible sample bias among those who reported years of math taken, a number of the female adjusted means are higher than the male adjusted means in the low and middle ability groups. This data suggests that equating for years of math may be a major factor in eliminating sex differences in performance in the low and middle ability ranges.

The only remaining question to be asked is whether number of math courses taken can be automatically equated with math background. Casserly (1983) has noted that females, especially those from lower socioeconomic classes, tend to take business math courses such as accounting and bookkeeping more frequently than males. It is possible, then, that males and females are reporting qualitatively different training when they report equal number of years of math, with females actually reporting a lower level of training.

Finally, there are many reasons why we would expect males to participate more effectively in academic activities, particularly in mathematics. As pointed out in the introductory section of this paper: 1) females receive less praise for correct answers in math than males; 2) males are praised for participation in academic activities more frequently than females; and 3) teachers sex-stereotype academic fields making more academic contacts with females in reading and with males in math.

Another factor might be different out-of-class experiences which influence "baseline" ability of males and females by the time they begin high school math courses. These out-of-class experiences include taking mechanical items apart, participating in strategy-memory game playing and playing geometrical or trigonometric sports (sailing, billiards, etc.).

While this study shows large differences between males and females in the high ability groups, it is important to note that there are relatively few students of either sex in the high ability groups. Only 13% of the 1978 students and 11% of the 1979 students fall into these groups. Except for geometry, almost all of the scale differences between males and females are nonsignificant in the lower and middle ability groups which include the majority of the population.

This study indicates that the much publicized superiority of males in mathematics applies to the relatively small percentage of males in the high ability group. It is important to note that this group is probably encouraged more in mathematics than high ability females or either males or females in any of the other ability groups.

TABLE 1*

1978 Subscales

Sex Neutral

	Algebra	Geometry	Arithmetic	Miscellaneous
Regular Math	8	11	13	3
Quantitative Comparison	6	6	6	1

Male Content

	Algebra	Geometry	Arithmetic	Miscellaneous
Regular Math	0	1	0	2
Quantitative Comparison	0	0	0	0

Female Content

	Algebra	Geometry	Arithmetic	Miscellaneous
Regular Math	2	0	0	0
Quantitative Comparison	0	0	0	0

* Items total to 59 instead of 60 because one item was not used in this administration

TABLE 2

1979 Subscales

Sex Neutral

	Algebra	Geometry	Arithmetic	Miscellaneous
Regular Math	9	11	10	5
Quantitative Comparison	5	6	5	2

Male Content

	Algebra	Geometry	Arithmetic	Miscellaneous
Regular Math	2	0	2	1
Quantitative Comparison	1	0	0	0

Female Content

	Algebra	Geometry	Arithmetic	Miscellaneous
Regular Math	0	0	0	0
Quantitative Comparison	0	0	1	0

TABLE 3

Male and Female Content Items - 1978

Male-Content

3. A farmer has completed fencing all but one side of a square field. If he has already used 3x meters of fencing, how many meters will he need for the last side.

14. Tri-School Event

	Event I	Event II	Event III
First Place (5 points)	C	A	
Second Place (3 points)	A		
Third Place (1 point)	B		

In a three school meet, schools A, B, and C each entered one team for each of 3 events. If the score card above is completed and shows no ties in any event, what is the greatest possible number of points by which B's total score could exceed A's total score?

21. Five contestants competed for two days in a sports event in which there were no ties in standing . At the end of each day, the scores were recorded. Which of the following could NOT be the total number of contestants whose standings changed on the second day?

Female-Content

53. A kindergarten class wants to buy a $77 tropical tree for the school. If the teacher agrees to pay twice as much as the class, and the administration promises to pay 4 times as much as the class, how much should the teacher pay?

57. Starting at A, Julia traveled 2/5 of the distance from A to B by bus, then traveled 15 miles by car, and then walked 3 miles to reach B. For what fractional part of her trip from A to B was Julia traveling by bus or car?

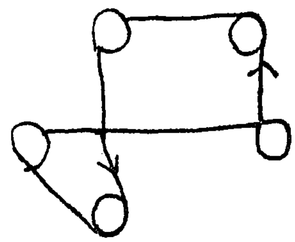
TABLE 4

Male And Female Items - 1979

Male - Content

17. Jack begins reading at the top of page N and finishes at the bottom of page R. If the pages are numbered and read consecutively and if there are no blank pages, how many pages has he read?

18. Out of a total of 154 games played, a ball team won 54 more games than it lost. If there were no ties, how many games did they win?

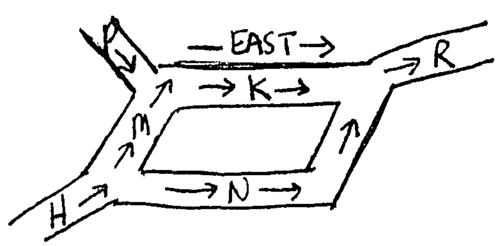


28. In the figure above, a belt which runs over five wheels moves in the direction of the arrows. How many of the wheels are turning clockwise?

Quantitative Comparison (asks student to determine which quantity is larger)

40. The number of articles X can producer per hour if he produces at the rate of 25 articles in 15 minutes	The number of articles Y can producer per hour if he produces at the rate of 20 articles in 12 minutes.

57. A person is paid time and a half for all hours worked in excess of 7.5 hours per day. If the person works 10 hours in one day, by what percent are his regular wages for the day increased?



60. In the highway system illustrated above, one-tenth of the eastbound cars from highway H turn into M; the rest continue on N. One fifth of the eastbound cars on K come from P; the rest, from M. If the traffic on all highways in is the direction of the arrows, what is the ratio of the traffic on H to that on R.

TABLE 4 (continued)

Female - Content

Quantitative Comparison (asks student to determine which quantity is larger)

51. Beef Prices

Brand	Price Per Pound
X	$1.40
Y	$1.20

Average price per pound of 2 pounds of brand X beef and 3 pounds of brand Y beef	Average price per pound of 2 pounds of brand X beef and 1 pound of brand Y beef

TABLE 5

1978 Miscellaneous Math

```
          18     6
      ¶‾‾‾‾‾‾‾‾‾‾‾‾‾¶
   12 ¶  .       .  ¶
      ¶             ¶
    . ¶  .       3  ¶
      ¶_____________¶
```

8. In the figure above, each dot represents a missing number. The product of numbers in any row or column inside the square must be equal to the number in the corresponding position outside the square. If the table is completed, which number would go below?

14.

	Tri-School Meet		
	Event I	Event II	Event III
First Place (5 points)	C	A	
Second Place (3 points)	A		
Third Place (1 point)	B		

In a three-school meet, Schools A, B, and C, each entered one team for each of 3 events. If the score care above is completed and shows no ties in any event, what is the greatest possible number of points by which B's total score could exceed A's total score?

21. Five contestants competed for two days in a sports event in which there were no ties in standings. At the end of each day, the standings were recorded. Which of the follwoing could NOT be the total number of contestants whose standings changed on the second day.

30. If 7! means 7 6 5 4 3 2 1
$7!_e$ means 6 4 2
and $7!_o$ means 7 5 3 1
which of the following is true?

45. Quantitative Comparison - asks which of the two quantities is greater:

◇n = $(n - 4)^2$ for all positive integers n

◇3 ◇5

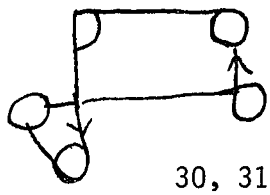
55. If m (1/n) = 1, which of the following is NOT necessarily true?

TABLE 6

1979 Miscellaneous Math

15. If the pattern of the first 5 arrows above continues to the right, the 107th arrow would be in which of the following directions?

24. $\begin{array}{r} AB \\ \underline{CD} \\ AAA \end{array}$ In the correct addition problem shown, A, B, C, and D represent nonzero digits. What is the value of C?

28. 

In the figure shown, a belt which runs over five wheels moves in the direction of the arrows. How many of the wheels turn clockwise?

30. 30, 31, 32, 33, 34, 35

If S is the sum of the remainders when each number in the list is divided by 6, what is the remainder when S is divided by 6?

For questions 54 and 55 refer to the following definition: For all real numbers p and q, where $q \neq 0$. $p \searrow_q = \frac{p^2}{q^2}$

54. $\left(3 \searrow_5\right)\left(5 \searrow_6\right) =$

55. If pqrs = 0, which of the following is(are) necessarily true?

I. $p \searrow_q = q \searrow_p$ II. $\left(p \searrow_q\right)\left(1/p \searrow_{1/q}\right) = 1$ III. $\left(p \searrow_q\right)\left(r \searrow_s\right) = pr \searrow_{qs}$

Quantitative Comparison - asks student to determine which of two quantities is larger

37. In a certain class, every student studies art or science of both; 12 study only art, 10 study only science and X study both.

Number of students in the class who study art but not science	Number of students in the class who study science but not art

49. In a game a fair coin is to be tossed 10 times and the score will be the number of heads obtained

The chances of a score of exactly seven	The chances of a score of exactly three

TABLE 7

Means and Standard Deviations of Males and Females on Subscales
1978 SAT-M

	Males	Females	Significance
Regular Math/Sex Neutral			
Algebra (n = 8)	4.16 (s=2.00)	4.06 (s=1.97)	.17
Geometry (n = 11)	4.79 (s=2.40)	4.49 (s=2.29)	.0004
Arithmetic (n = 13)	7.04 (s=2.89)	6.83 (s=2.90)	.05
Miscellaneous (n = 3)	1.65 (s=1.09)	1.76 (s=1.03)	.0036
Quantitative Comparison/Sex Neutral			
Algebra (n = 6)	3.93 (s=1.67)	3.30 (s=1.70)	.12
Geometry (n = 6)	3.15 (s=1.60)	2.94 (s=1.60)	.0005
Arithmetic (n = 6)	3.51 (s=1.40)	3.42 (s=1.34)	.097
Miscellaneous (n = 1)	.41 (s= .49)	.41 (s= .49)	.77
Regular Math/Male Content			
Geometry (n = 1)	.81 (s= .39)	.80 (s= .40)	.31
Miscellaneous (n = 2)	.98 (s= .39)	.85 (s= .73)	.00001
Regular Math/Female Content			
Algebra (n = 2)	.85 (s= .76)	.80 (s= .73)	.07

TABLE 8

Means and Standard Deviations of Males and Females on Subscales
1979 SAT-M

	Males	Females	Significance
Regular Math/Sex Neutral			
Algebra (n = 9)	5.18 (s=2.36)	5.16 (s=2.26)	.81
Geometry (n = 11)	4.86 (s=2.50)	4.59 (s=2.32)	.002
Arithmetic (n = 10)	6.15 (s=2.45)	6.04 (s=2.45)	.24
Miscellaneous (n = 5)	2.28 (s=1.41)	1.36 (s=1.36)	.44
Quantitative Comparison/Sex Neutral			
Algebra (n = 5)	2.99 (s=1.50)	2.92 (s=1.48)	.22
Geometry (n = 6)	3.65 (s=1.70)	3.57 (s=1.48)	.22
Arithmetic (n = 5)	3.26 (s=1.30)	3.23 (s=1.36)	.41
Miscellaneous (n = 2)	1.08 (s= .73)	1.08 (s= .71)	.76
Regular Math/Math Content			
Algebra (n = 2)	.62 (s= .62)	.51 (s= .56)	.0000
Arithmetic (n = 2)	.56 (s= .65)	.57 (s= .64)	.57
Miscellaneous (n = 1)	.88 (s= .32)	.83 (s= .37)	.00001
Quantitative Comparison/Male Content			
Algebra (n = 1)	.64 (s= .48)	.63 (s= .48)	.70
Quantitative Comparison/Female Content			
Arithmetic (n = 1)	.41 (s= .49)	.38 (s = .49)	.13

TABLE 9

MEANS AND STANDARD DEVIATIONS OF COMBINED 1978 SCALES

	Males	Females	Significance
Algebra n = 16	8.4 (s=3.87)	8.15 (s=3.82)	.10
Geometry n = 18	8.75 (s=3.49)	8.23 (s=3.76)	.001
Arithmetic n = 19	10.55 (s=3.9)	10.25 (s=3.89)	.05
Miscellaneous n = 6	3.03 (s=1.82)	3.02 (s=1.71)	.70
Regular Math n = 40	20.28 (s=8.6)	19.58 (s=8.3)	.05
Quantitative Comparisons n = 20	10.46 (s=4.18)	10.08 (s=4.19)	.05
Male Content n = 3	1.79 (s= .92)	1.64 (s=.92)	.001
Female Content n = 2	.84 (s= .75)	.70 (s= .73)	.10

TABLE 10

MEANS AND STANDARD DEVIATIONS OF

COMBINED 1979 SCALES

	Males	Females	Significance
Algebra (n = 17)	9.43 (s = 4.17)	9.23 (s = 4.04)	.08
Geometry (n = 17)	8.51 (s = 3.89)	8.16 (s = 3.73)	.01
Arithmetic (n = 18)	10.38 (s = 4.05)	10.23 (s = 4.06)	.28
Miscellaneous (n = 8)	4.25 (s = 1.95)	4.15 (s = 1.90)	.17
Regular Math (n = 40)	20.54 (s = 8.47)	19.96 (s = 8.07)	.05
Quantitative Comparisons (n = 20)	12.04 (s = 4.90)	11.81 (s = 4.91)	.20
Male Content (n = 4)	2.70 (s = 1.41)	2.54 (s = 1.38)	.002
Female Content (n = 1)	.41 (s = .49)	.38 (s = .49)	.13

TABLE 11

SUBSCALE RELIABILITIES

	1978					1979			
	Male		Female			Male		Female	
	n*	15	n	15		n	15	n	15
Regular Math/ Sex Neutral									
Algebra n = 8	.71	.82	.70	.80	n=9	.75	.83	.73	.82
Geometry n = 11	.69	.75	.66	.73	n=11	.74	.79	.70	.76
Arithmetic n = 13	.75	.77	.76	.78	n=10	.60	.82	.57	.80
Miscellaneous n = 3	.52	.84	.56	.86	n=5	.60	.82	.57	.80
Quantitative Comparison/Sex Neutral									
Algebra n = 6	.60	.83	.62	.85	n=5	.64	.84	.63	.84
Geometry n = 6	.60	.79	.61	.80	n=6	.68	.84	.70	.85
Arithmetic n = 6	.51	.76	.47	.69	n=5	.55	.79	.56	.79
Miscellaneous** n = 1					n=2	.31	.77	.27	.74
Regular Math/Male Content									
Arithmetic					n=2	.21	.68	.23	.68
Algebra					n=2	.25	.71	.05	.28
Miscellaneous n=2	.37	.82	.30	.76					
Regular Math/Female Content									
Algebra n=2	.43	.84	.40	.83					

* Columns labeled "n" show reliabilities for number of items in scale. Columns labeled 15 show reliabilities adjusted for 15 item scales using Spearman-Brown formula.

** Subscales with one item are not included.

TABLE 12

SIGNIFICANT PARTIAL PHI COEFFICIENTS

1978

Male				Female			
Item #	ø	Classif.	Δ	Item #	ø	Classif.	Δ
50	.111	Geom QC	15.9	30	.1	Misc RM	11.9
14	.073	Misc RM Male Content	11.2	31	.071	Arith RM	10.1
20	.072	Alg RM	17.5	47	.07	Alg QC	14.1
4	.067	Arith RM	8.2	8	.069	Misc RM	19.7
17	.062	Geom RM	14.8	9	.067	Alg RM	10.6
46	.060	Geom QC	13.9	60	.043	Arith RM	17.4
44	.044	Alg QC	12.9	1	.041	Alg RM	6.8
23	.041	Arith RM	17.8	51	.037	Geom QC	16.8
56	.040	Alg RM	15.6	42	.031	Alg QC	11.8
21	.039	Misc RM Male Content	16.2	13	.031	Arith RM	11.8
25	.034	Alg RM	20.4				
27	.033	Arith RM	9.1				
6	.032	Arith RM	8.9				

TABLE 13

SIGNIFICANT PARTIAL PHI COEFFICIENTS

1979

Male				Female			
Item #	Ø	Classif.	Δ	Item #	Ø	Classif.	Δ
21	.089	Geom RM	16.2	17	.080	Arith RM Male Content	13.3
60	.080	Alg RM Male Content	7.1	54	.070	Misc RM	12.6
23	.080	Geom RM	17.1	10	.069	Alg RM	10.4
18	.077	Alg RM Male Content	13.1	47	.068	Arith QC	13.4
28	.068	Misc RM Male Content	7.9	12	.050	Alg RM	11.7
46	.065	Geom QC	13.1	45	.044	Geom QC	12.3
24	.060	Misc RM	19.0	7	.043	Alg RM	9.3
57	.060	Arith RM	17.2	37	.036	Misc QC	9.7
32	.057	Alg RM	12.1	26	.035	Arith RM	7.8
36	.053	Arith QC	9.9	4	.035	Alg RM	8.3
14	.046	Arith RM	11.8				
25	.044	Geom RM	17.5				
31	.040	Arith RM	12.0				
33	.033	Arith QC	6.6				

TABLE 14

SIGNIFICANT PARTIAL PHI COEFFICIENTS

BY SCALE

MALES - 1978

SCALE	Number	% of Scale	NON-SCALE	Number	% of Non-Scale
Algebra	4	25%	Non-Algebra	9	20%
Geometry	3	17%	Non-Geometry	10	24%
Arithmetic	4	21%	Non-Arithmetic	9	23%
Miscellaneous	2	33%	Non-Miscellaneous	11	20%
Regular Math	10	25%	Non-Regular Math (Quantitative Comparison)	3	15%
Quantitative Comparison	3	15%	Non-Quantitative Comparison (Regular Math)	10	25%
Male Content	2	67%	Non-Male Content	11	20%
Female Content	0	0	Non-Female Content	13	23%

Non-scale means all items remaining on test which are not included in scale indicated in previous column

TABLE 15

SIGNIFICANT PARTIAL PHI COEFFICIENTS

BY SCALE

FEMALES - 1978

SCALE	Number	% of Scale	NON-SCALE	Number	% of Non-Scale
Algebra	4	25%	Non-Algebra	6	13%
Geometry	1	5%	Non-Geometry	9	21%
Arithmetic	3	16%	Non-Arithmetic	7	18%
Miscellaneous	2	33%	Non-Miscellaneous	8	15%
Regular Math	7	17%	Non-Regular Math (Quantitative Comparison)	3	15%
Quantitative Comparison	3	15%	Non-Quantitative Comparison (Regular Math)	7	17%
Male Content	0		Non-Male Content	10	18%
Female Content	0	0	Non-Female Content	10	18%

Non-scale means all items remaining on test which are not included in scale indicated in previous column

TABLE 16

SIGNIFICANT PARTIAL PHI COEFFICIENTS

BY SCALE

MALES - 1979

SCALE	Number	% of Scale	NON-SCALE	Number	% of Non-Scale
Algebra	3	18%	Non-Algebra	11	26%
Geometry	4	24%	Non-Geometry	10	22%
Arithmetic	5	28%	Non-Arithmetic	9	21%
Miscellaneous	2	25%	Non-Miscellaneous	12	23%
Regular Math	11	28%	Non-Regular Math (Quantitative Comparison)	3	15%
Quantitative Comparison	3	15%	Non-Quantitative Comparison	11	28%
Male Content	4	67%	Non-Male Content	11	19%
Female Content	0	0	Non-Female Content	14	24%

Non-scale means all items remaining on test which are not included in scale indicated in previous column

TABLE 17

SIGNIFICANT PARTIAL PHI COEFFICIENTS

BY SCALE

FEMALES - 1979

SCALE	Number	% of Scale	NON-SCALE	Number	% of Non-Scale
Algebra	4	24%	Non-Algebra	7	16%
Geometry	1	6 %	Non-Geometry	10	23%
Arithmetic	4	22%	Non-Arithmetic	7	17%
Miscellaneous	2	25%	Non-Miscellaneous	9	17%
Regular Math	8	20%	Non-Regular Math (Quantitative Comparison)	3	15%
Quantitative Comparison	3	15%	Non-Quantitative Comparison (Regular Math)	8	20%
Male Content	1	17%	Non-Male Content	10	19%
Female Content	0	0	Non-Female Content	11	19%

Non-scale means all items remaining on test which are not included in scale indicated in previous column

TABLE 18

SUMMARY OF RESULTS

HYPOTHESES 1 - 7

Hypotheses	Significant Partial Phi Coefficients
1. Male Content	
1978	Confirmed
1979	Confirmed
2. Geometry	
1978	Partially Confirmed
1979	Confirmed
3. Regular Math	
1978	Confirmed
1979	Confirmed
4. Algebra	
1978	Confirmed
1979	Confirmed
5. Miscellaneous	
1978	Confirmed
1979	Confirmed
6. Quantitative Comparison	
1978	Not Confirmed
1979	Not Confirmed
7. Female Content*	
1978	Not Confirmed
1979	Not Confirmed

* Female Content scale has too few items to draw conclusions.

TABLE 19

MEANS OF SELECTED SAMPLE OF

MALES AND FEMALES BY SDQ7*

Years of Math	1978 Males x	1978 Males n	1978 Females x	1978 Females n	1979 Males x	1979 Males n	1979 Females x	1979 Females n
1	341	26	330	37	338	20	341	42
2	383	130	378	192	366	121	368	162
3	434	232	438	317	422	213	444	303
4	514	436	520	374	516	462	519	430
5	582	120	585	103	569	124	577	90
Did Not Take Math	403	46	392	65	430	59	357	86
No Answer	489	511	507	412	507	501	507	387

*Means are for sample of 3,000 (1,500 males and 1,500 females) from 1978 and 1979 administration. Within each sample of 3,000, 1,000 candidates (500 males and 500 females) were selected with SAT-M scores of 200-399; 1,000 candidates (500 males and 500 females) were selected with SAT-M scores of 400-599 and 1,000 candidates (500 males and 500 females) were selected with SAT-M scores of 600 - 800.

TABLE 20

SIGNIFICANCE OF COVARIANCE " F" TEST FOR SEX

CONTROLLING FOR YEARS OF MATH

1978	200 - 399	400 - 599	600 - 800
Scaled Math	.025	.37	.0001
Algebra	.10	.67	.02
Geometry	.0006	.026	.0001
Arithmetic	.056	.37	.002
Miscellaneous	.48	.41	.18
Quantitative Comparison	.002	.03	.04
Regular Math	.04	.67	.0001
Male Content	.002	.11	.03
Female Content	.009	.62	.04
1979			
Scaled Math	.77	.91	.001
Algebra	.89	.92	.0001
Geometry	.93	.03	.02
Arithmetic	.84	.77	.53
Miscellaneous	.13	.90	.02
Quantitative Comparison	.28	1.00	.03
Regular Math	.66	.18	.001
Male Content	.07	.09	.035
Female Content	.63	.84	.11

TABLE 21A

1978

ADJUSTED MEANS OF COVARIANCE

FOR SEX CONTROLLING FOR YEARS OF MATH

	200-399		400-599		600-800	
	Males	Females	Males	Females	Males	Females
Scaled Math	325.10	317.42	485.54	481.79	654.52	639.30
Algebra	4.31	4.08	8.21	8.14	12.56	12.27
Geometry	4.86	4.37	8.53	8.16	12.85	12.13
Arithmetic	6.36	6.05	10.58	10.45	14.66	14.21
Miscellaneous	1.32	1.38	2.90	2.98	4.83	4.72
Regular Math	10.73	10.21	19.99	19.87	29.89	28.63
Quantitative Comparison	6.13	5.67	10.24	9.87	15.00	14.70
Male Content	1.03	.86	1.82	1.73	2.48	2.37
Female Content	.03	.02	.77	.79	1.48	1.38

TABLE 21B

1979

ADJUSTED MEANS OF COVARIANCE

FOR SEX CONTROLLING FOR YEARS OF MATH

	200-399		400-599		600-800	
	Males	Females	Males	Females	Males	Females
Scaled Math	326.00	325.02	484.92	484.45	644.18	634.38
Algebra	4.66	4.68	9.62	9.63	13.74	13.28
Geometry	4.48	4.49	8.17	7.82	12.60	12.24
Arithmetic	5.73	5.76	10.45	10.40	14.62	14.54
Miscellaneous	2.57	2.44	3.92	3.93	6.25	6.05
Regular Math	10.96	11.07	20.12	19.75	29.90	29.08
Quantitative Comparison	6.49	6.31	12.03	12.03	17.32	17.32
Male Content	1.45	1.33	2.63	2.49	3.94	3.79
Female Content	.11	.10	.33	.34	.78	.72

TABLE 22

CORRELATION OF HIGH SCHOOL GRADES WITH

SAT-M AND SCALE SCORES

	1978		1979	
	Male	Female	Male	Female
SAT-M	.49	.49	.51	.47
Algebra	.45	.45	.49	.43
Geometry	.44	.43	.47	.44
Arithmetic	.49	.48	.47	.44
Miscellaneous	.45	.38	.44	.39
Regular Math	.49	.48	.50	.44
Quantitative Comparison	.45	.45	.50	.45
Male Content	.26	.32	.40	.36
Female Content	.31	.30	.39	.27

TABLE 23

CORRELATIONS OF SUBSCALES WITH SAT-M

	1978			1979		
		Male	Female		Male	Female
Regular Math/Sex Neutral						
Algebra	n=8	.85	.82	n=9	.85	.84
Geometry	n=11	.80	.79	n=11	.76	.71
Arithmetic	n=13	.87	.85	n=10	.85	.84
Miscellaneous	n=3	.73	.70	n=5	.74	.73
Quantitative Comparison/Sex Neutral						
Algebra	n=6	.76	.77	n=5	.77	.78
Geometry	n=6	.77	.78	n=6	.82	.82
Arithmetic	n=6	.70	.70	n=5	.73	.73
Miscellaneous	n=1	.51	.45	n=2	.56	.54
Regular Math/Male Content						
Arithmetic				n=2	.52	.48
Algebra				n=2	.62	.59
Miscellaneous	n=2	.66	.58	n=1	.21	.33
Geometry	n=1	.41	.49			
Quantitative Comparison/Male Content						
Algebra				n=1	.56	.56
Regular Math/Female Content						
Algebra	n=2	.63	.65			
Quantitative Comparison/Female Content						
Arithmetic				n=1	.53	.51

TABLE 24

CORRELATIONS OF VERBAL SCALES WITH MATH SCALES - MALES - 1978

	SAT-M	Alg.	Geom.	Arith	Misc.	Reg.Math	Quan.Com	Male Content	Female Content
SAT-V n = 85	.71	.67	.62	.66	.62	.69	.67	.52	.47
Vocab n = 45	.65	.63	.60	.61	.58	.65	.62	.49	.45
Reading n = 40	.67	.65	.60	.62	.60	.66	.64	.51	.46
Antonym n = 25	.59	.57	.54	.55	.51	.596	.56	.43	.40
Sent. Com. n = 15	.64	.61	.55	.60	.57	.62	.59	.47	.43
Read Comp. n = 25	.62	.60	.56	.57	.55	.61	.59	.47	.43
Analogies n = 20	.64	.62	.58	.60	.58	.64	.61	.48	.45

TABLE 25

CORRELATIONS OF VERBAL SCALES WITH MATH SCALES - FEMALES - 1978

	SAT-M	Alg.	Geom.	Arith	Misc.	Reg.Math	Quan.Com	Male Content	Female Content
SAT-V n = 85	.76	.71	.66	.71	.65	.74	.69	.58	.53
Vocab n = 45	.70	.66	.61	.67	.62	.70	.65	.54	.50
Reading n = 40	.63	.69	.65	.68	.64	.73	.68	.57	.53
Antonym n = 25	.64	.61	.56	.61	.56	.64	.59	.49	.46
Sent. Com. n = 15	.69	.65	.60	.64	.690	.67	.64	.56	.49
Read Comp. n = 25	.69	.65	.62	.64	.60	.69	.63	.52	.50
Analogies n = 20	.69	.65	.60	.65	.62	.68	.64	.54	.48

TABLE 26

CORRELATIONS OF VERBAL SCALES WITH MATH SCALES - MALES - 1979

	SAT-M	Alg.	Geom.	Arith	Misc.	Reg.Math	Quan.Com	Male Content	Female Content
SAT-V n = 85	.75	.69	.70	.70	.60	.73	.70	.60	39
Vocab n = 45	.68	.63	.65	.65	.55	.68	.63	.54	.35
Reading n = 40	.72	.67	.68	.68	.59	.72	.67	.59	.38
Antonym n = 25	.59	.55	.57	.57	.49	.60	.55	.47	.29
Sent. Com. n = 15	.64	.58	.62	.60.	.52	.63	.60	.52	.35
Read Comp. n = 25	.70	.65	.65	.66	.58	.69	.65	.57	.37
Analogies n = 20	.66	.61	.625	.63	.54	.65	.62	.54	.34

TABLE 27

CORRELATIONS OF VERBAL SCALES WITH MATH SCALES - FEMALES - 1979

	SAT-M	Alg.	Geom.	Arith	Misc.	Reg.Math	Quan.Com	Male Content	Female Content
SAT-V n = 85	.74	.67	.66	.69	.64	.72	.70	.57	.40
Vocab n = 45	.66	.60	.60	.63	..59	.66	.63	.52	.35
Reading n = 40	.72	.66	.66	.68	.64	.70	.69	.57	.41
Antonym n = 25	.59	.55	.54	.56	.53	.59	.56	.46	.31
Sent.Com. n = 15	.65	.60	.60	.62	.58	.64	.62	.52	.37
Read Comp. n = 25	.70	.65	.64	.66	.62	.69	.67	.55	.41
Analogies n = 20	.65	.59	.60	.63	.59	.65	.62	.52	.35

TABLE 28

PREDICTED MEANS

MALES 1978

	Items in Original Scale	Percent Passing*	Projected Scale Length 15	30	45	60
Regular Math/ Sex neutral						
Algebra	8	52%	7.8	15.60	23.20	31.20
Geometry	11	44%	6.6	13.20	19.80	26.40
Arithmetic	13	54%	8.10	16.20	24.30	32.40
Miscellaneous	3	55%	8.25	16.50	24.75	33.00
Quantitative Comparison/ Sex Neutral						
Algebra	6	57%	9.90	19.80	29.70	39.60
Geometry	6	53%	7.95	15.90	23.85	31.80
Arithmetic	6	58%	8.85	17.70	26.55	35.40
Miscellaneous	1	41%	6.15	12.30	18.45	24.60
Regular Math/ Male Content						
Geometry	1	81%	12.15	24.30	36.45	48.60
Miscellaneous	2	49%	7.35	14.70	22.05	29.40
Regular Math/ Female Content						
Algebra	2	43%	6.30	12.60	18.90	25.20

* Percent passing refers to the percent of scale items which males answered correctly

TABLE 29

PREDICTED MEANS

FEMALES 1978

	Items in Original Scale	Percent Correct*	Projected Scale Length 15	30	45	60
Regular Math/ Sex Neutral						
Algebra	8	51%	7.65	15.30	22.95	30.60
Geometry	11	41%	6.15	12.60	20.50	24.60
Arithmetic	13	52%	7.95	15.90	23.85	31.80
Miscellaneous	3	59%	8.85	17.70	26.55	35.40
Quantitative Comparison/ Sex Neutral						
Algebra	6	55%	8.25	16.50	24.75	33.00
Geometry	6	49%	7.35	14.70	22.05	29.40
Arithmetic	6	57%	8.55	17.10	25.65	34.20
Miscellaneous	1	41%	6.15	12.30	18.45	24.60
Regular Math/ Male Content						
Geometry	1	84%	12.00	24.00	36.00	48.00
Miscellaneous	1	80%	6.30	12.60	18.90	25.20
Regular Math/ Female Content						
Algebra	2	80%	6.00	12.00	18.00	24.00

* Percent correct refers to the percent of scale items which females answered correctly

TABLE 30

PREDICTED MEANS

MALES 1979

	Items in Original Scale	Percent Correct*	Projected Scale Length			
			15	30	45	60
Regular Math/ Sex Neutral						
Algebra	9	58%	9.75	19.50	29.25	39.00
Geometry	11	44%	6.60	13.20	19.80	26.40
Arithmetic	10	47%	9.30	18.60	27.90	37.20
Miscellaneous	5	46%	6.90	13.80	20.70	27.60
Quantitative Comparison/ Sex Neutral						
Algebra	5	60%	9.00	18.00	27.00	36.00
Geometry	6	61%	9.15	18.30	27.45	36.60
Arithmetic	5	65%	9.75	19.50	29.25	39.00
Miscellaneous	2	54%	8.10	16.20	24.30	32.40
Regular Math/ Male Content						
Arithmetic	2	28%	4.20	8.40	12.60	16.80
Algebra	2	31%	4.65	9.30	13.95	18.60
Miscellaneous	1	88%	13.20	26.40	39.60	52.80
Quantitative Comparison/ Male Content						
Algebra	1	61%	9.60	19.20	28.80	38.40
Quantitative Comparison/ Female Content						
Arithmetic	1	41%	6.15	12.30	18.45	24.60

* Percent correct refers to the percent of scale items which males answered correctly

TABLE 31

PREDICTED MEANS

FEMALES 1979

	Items in Original Scale	Percent Passing	Projected Scale Length 15	30	45	60
Regular Math/ Sex Neutral						
Algebra	9	57%	9.75	19.50	29.25	39.00
Geometry	11	42%	6.30	12.60	18.90	25.20
Arithmetic	10	60%	9.00	18.00	27.00	36.00
Miscellaneous	5	45%	4.05	8.10	12.15	16.20
Quantitative Comparison/ Sex Neutral						
Algebra	5	58%	8.85	17.70	26.55	35.40
Geometry	6	60%	9,00	18.00	27.00	36.00
Arithmetic	5	64%	9.75	19.50	29.25	39.00
Miscellaneous	2	54%	8.10	16.20	24.30	32.40
Regular Math/ Male Content						
Arithmetic	2	29%	4.35	8.70	13.05	17.40
Algebra	2	26%	3.90	7.80	11.70	15.60
Miscellaneous	1	83%	12.45	24.90	37.35	49.80
Quantitative Comparison/ Male Content						
Algebra	1	63%	9.45	18.90	28.35	37.80
Quantitative Comparison/ Female Content						
Arithmetic	1	38%	5.70	11.40	17.10	22.80

TABLE 32

PREDICTED VARIANCES

MALES 1978

	Items in Original Scale	Percent Passing*	Projected Scale Length 15	30	45	60
Regular Math/ Sex Neutral						
Algebra	8	52%	12.15	24.30	36.45	48.60
Geometry	11	44%	12.45	24.90	37.35	49.80
Arithmetic	13	54%	13.80	27.60	41.40	55.20
Miscellaneous	3	55%	6.75	13.50	20.25	27.00
Quantitative Comparison/ Sex Neutral						
Algebra	6	57%	8.40	16.80	25.20	33.60
Geometry	6	53%	8.10	16.20	24.30	32.60
Arithmetic	6	58%	5.40	10.80	16.20	21.60
Miscellaneous	1	41%	7.35	14.70	22.05	29.40
Regular Math/ Male Content						
Geometry	1	81%	5.85	11.70	17.55	23.40
Miscellaneous	2	49%	2.40	4.80	7.20	9.60
Regular Math/ Female Content						
Algebra	2	43%	3.15	6.30	9.45	12.60

* Percent correct refers to the percent of scale items which males answered correctly.

TABLE 33

PREDICTED VARIANCES

FEMALES 1978

	Items in Original Scale	Percent Passing*	Projected Scale Length 15	30	45	60
Regular Math/ Sex Neutral						
Algebra	8	51%	11.25	22.50	33.75	45.00
Geometry	11	41%	10.50	21.50	31.50	42.00
Arithmetic	13	52%	16.05	32.10	48.15	64.20
Miscellaneous	3	59%	5.25	10.50	15.75	21.00
Quantitative Comparison/ Sex Neutral						
Algebra	6	55%	8.93	17.85	26.75	35.70
Geometry	6	49%	8.25	16.50	24.75	33.00
Arithmetic	6	57%	4.80	9.60	14.40	19.20
Miscellaneous	1	41%	7.35	14.70	22.05	29.40
Regular Math/ Male Content						
Geometry	1	84%	6.00	12.00	18.00	24.00
Miscellaneous	1	80%	2.06	4.10	6.17	8.20
Regular Math/ Female Content						
Algebra	2	79%	2.70	5.50	8.24	10.98

* Percent correct refers to percent of items in scale which females answered correctly

TABLE 34

PREDICTED VARIANCES

MALES 1979

	Items in Original Scale	Percent Passing*	Projected Scale Length 15	30	45	60
Regular Math/ Sex Neutral						
Algebra	9	58%	15.30	30.60	45.90	61.20
Geometry	11	44%	14.25	28.50	42.75	57.00
Arithmetic	10	47%	14.70	29.40	44.10	58.80
Miscellaneous	5	46%	7.65	15.30	22.95	30.60
Quantitative Comparison/ Sex Neutral						
Algebra	5	60%	9.00	18.00	27.00	36.00
Geometry	6	61%	11.55	23.10	34.65	46.20
Arithmetic	5	65%	8.25	16.50	24.75	33.00
Miscellaneous	2	54%	1.46	2.91	4.37	5.82
Regular Math/ Male Content						
Arithmetic	2	28%	1.17	2.34	3.51	4.68
Algebra	2	31%	2.69	5.37	8.06	10.74
Miscellaneous	1	88%	4.80	9.60	14.40	19.20
Quantitative Comparison/ Male Content						
Algebra	1	64%	7.20	14.40	21.60	28.80
Quantitative Comparison/ Female Content						
Arithmetic	1	41%	7.35	14.70	22.05	29.40

* Percent correct refers to percent of items in scale which females answered correctly

TABLE 35

PREDICTED VARIANCES

FEMALES 1979

	Items in Original Scale	Percent Passing*	Projected Scale Length 15	30	45	60
Regular Math/ Sex Neutral						
Algebra	9	57%	9.15	18.30	27.45	36.60
Geometry	11	42%	12.30	24.60	36.90	49.20
Arithmetic	10	60%	10.20	20.40	30.60	40.80
Miscellaneous	5	45%	6.75	13.50	20.25	27.00
Quantitative Comparison/ Sex Neutral						
Algebra	5	58%	8.55	17.10	25.65	34.20
Geometry	6	60%	11.10	22.20	33.30	44.40
Arithmetic	5	64%	6.60	13.20	19.80	26.40
Miscellaneous	2	54%	1.80	3.60	5.40	7.20
Regular Math/ Male Content						
Arithmetic	2	29%	1.35	2.70	4.05	5.40
Algebra	2	26%	.30	.60	.90	1.20
Miscellaneous	1	83%	5.55	11.10	16.65	22.20
Quantitative Comparison/ Male Content						
Algebra	1	63%	7.20	14.40	21.60	28.80
Quantitative Comparison/ Female Content						
Arithmetic	1	38%	7.35	14.70	22.05	29.40

* Percent correct refers to percent of items in scale which females answered correctly

TABLE 36

PREDICTED RELIABILITIES

MALES 1978

	Items in Original Scale	Percent Correct*	Projected Scale Length 15	30	45	60
Regular Math/ Sex Neutral						
Algebra	8	52%	.70	.73	.93	.95
Geometry	11	44%	.75	.86	.90	.92
Arithmetic	13	54%	.77	.88	.91	.94
Miscellaneous	5	45%	.84	.92	.94	.96
Quantitative Comparison/ Sex Neutral						
Algebra	6	57%	.83	.88	.92	.94
Geometry	6	53%	.79	.88	.92	.94
Arithmetic	6	58%	.76	.90	.93	.95
Miscellaneous	1	41%				
Regular Math/ Male Content						
Geometry	1	81%				
Miscellaneous	2	49%	.82	.90	.93	.95
Regular Math/ Female Content						
Algebra	2	43%	.84	.92	.94	.96

* Percent correct refers to percent of items in scale which males answered correctly

TABLE 37

PREDICTED RELIABILITIES

FEMALES 1978

	Items in Original Scale	Percent Correct*	Projected Scale Length 15	30	45	60
Regular Math/ Sex Neutral						
Algebra	8	51%	.80	.90	.93	.95
Geometry	11	41%	.73	.84	.89	.90
Arithmetic	13	52%	.78	.88	.92	.94
Miscellaneous	3	59%	.86	.93	.95	.96
Quantitative Comparison/ Sex Neutral						
Algebra	6	55%	.85	.89	.92	.94
Geometry	6	49%	.80	.89	.93	.94
Arithmetic	6	57%	.69	.82	.87	.90
Miscellaneous	1	41%				
Regular Math/ Male Content						
Geometry	1	84%				
Miscellaneous	2	80%	.76	.87	.91	.93
Regular Math/ Female Content						
Algebra	2	80%	.83	.90	.95	.97

* Percent correct refers to percent of items in scale which females answered correctly

TABLE 38

PREDICTED RELIABILITIES

MALES 1979

	Items in Original Scale	Percent Correct	Projected Scale Length 15	30	45	60
Regular Math/ Sex Neutral						
Algebra	9	58%	.83	.90	.94	.95
Geometry	11	44%	.79	.88	.92	.94
Arithmetic	10	47%	.81	.90	.93	.95
Miscellaneous	5	46%	.82	.90	.93	.95
Quantitative Comparison/ Sex Neutral						
Algebra	5	60%	.84	.91	.94	.96
Geometry	6	61%	.84	.91	.94	.96
Arithmetic	5	65%	.79	.88	.91	.94
Miscellaneous	2	54%	.77	.87	.91	.93
Regular Math/ Male Content						
Arithmetic	2	28%	.68	.80	.86	.89
Algebra	2	31%	.71	.83	.88	.91
Miscellaneous	1	80%				
Quantitative Comparison/ Male Content						
Algebra	1	61%				
Quantitative Comparison/ Female Content						
Arithmetic	1	41%				

TABLE 39

PREDICTED RELIABILITIES

FEMALES 1979

	Items in Original Scale	Percent Correct	Projected Scale Length 15	30	45	60
Regular Math/ Sex Neutral						
Algebra	9	57%	.82	.90	.93	.96
Geometry	11	42%	.76	.87	.90	.99
Arithmetic	10	60%	.80	.89	.93	.94
Miscellaneous	5	45%	.80	.88	.93	.94
Quantitative Comparison/ Sex Neutral						
Algebra	5	58%	.84	.91	.94	.95
Geometry	6	60%	.85	.92	.95	.96
Arithmetic	5	64%	.79	.88	.92	.94
Miscellaneous	2	54%	.74	.85	.89	.92
Regular Math/ Male Content						
Arithmetic	2	29%	.68	.80	.87	.90
Algebra	2	26%	.28	.44	.54	.61
Miscellaneous	1	83%				
Quantitative Comparison/ Male Content						
Algebra	1	63%				
Quantitative Comparison/ Female Content						
Arithmetic	1	38%				

* Percent correct refers to the percent of items which females answered correctly

REFERENCES

Aiken, Lewis R. Jr. Non-intellective variables and mathematics achievement: Directions for research. Journal of Social Psychology, 1970a, 8 (1), 28 - 36.

Aiken, Lewis R. Jr. Attitudes toward mathematics, Review of Educational Research, 1970b, 40, (4), 557 - 596.

Aiken, Lewis R. Jr. Research on attitudes towards mathematics, The Arithmetic Teacher, 1972, 19, (3), 229 - 34.

Aiken, Lewis R. Jr. Some speculations and findings concerning sex differences in mathematical abilities and attitudes. In Fennema Elizabeth (Ed.), Mathematics Learning: What Research Says About Sex Differences, December, 1975, Columbus, Ohio: ERIC Mathematics Science and Environmental Education Clearinghouse.

Aiken, Lewis R. Jr. Update on attitudes and other effective variables in learning mathematics. Review of Educational Research, 1976, 46, (2) 293 - 311.

Aiken, Lewis R. Jr. and Dreger, Ralph Mason, The effects of attitudes on performance in mathematics. Journal of Educational Psychology, 1961, 52, (1), 19 - 24.

Astin, Helen S. Sex differences in mathematical and scientific precocity. In J.C. Stanley, D.P. Keating and L.H. Fox (Eds), Mathematical Talent: Discovery, description and development, Baltimore: The Johns Hopkins University Press, 1974a, 70 - 86.

Astin, Helen S., Overview of the findings. In H. Astin, H. Suniewick and S. Dweck, Women: A Bibliography on Their Education and Careers. New York: Behavioral Publications Incorporated,1974b, 1 - 10.

Becker, Joanne Rossi, Differential treatment of females and males in mathematics classes, Journal for Research in Mathematics Education; , V12,N1,P40 - 53, Jan. 1981.

Benbow, Camilla and Stanley, Julian, Sex differences in mathematics ability: Fact or fiction, Science, Vol. 210, Dec. 1980.

Bergman, Jane, Are little girls being harmed by "Sesame Street"? In Judith Stacey, Susan Bereand and Joan Daniels (Eds) And Jill Came Tumbling After: Sexism in American Education, New York: Dell Publishing Company, Inc., 1974, 110 - 15.

Block, Jeanne Humphrey, Concpetions of sex role: Some cross-cultural and longitudinal perspectives. American Psychologist, 1973, 28, (6), 512 - 26.

Bock, R.D., "A family study of spatial ability", American Psychologist, 1967, 22, 571 (Abstract).

Bock, R.D., "Word and image: Sources of the verbal and spatial factor on mental test scores," Psychometrika, 38, December, 1973.

Bock, R.D. and Kolakowski, D.F., Further evidence of sex-linked major gene influence on human spatial visualizing ability, American Journal of Human Genetics, 1973, 25, 1 - 14.

Broverman, D.M. and Klaiber, E.L., Negative relationships between abilities, Psychometrika, 1969, 34, 5 - 20.

Buffery, A.W.H. and Gray, J.A., Sex differences in the development spatial and linguistic skills. In C. Ounsted and P.C. Taylor (Eds) Gender Differences: Their Ontogeny and Significance, Baltimore: Williams and Wilkins, 1972.

Burnett, Sarah, Spatial visualization and sex differences in quantitative ability, Intelligence, Oct-Dec., 1979.

Casserly, Patricia Lund, An assessment of factors affecting female participation in advanced placement in programs in mathematics, chemistry and physics. Report to the National Science Foundation, July, 1975.

Casserly, Patricia Lund, Personal communication, 1983.

Christman, John F., Vidulick, Robert N., Drolle, Penelope and Kirk, McAnn, Factors influencing the selection of science careers by women, Paper presented at the symposium of the Committee on Women and Physics, 1976, Spring Meeting of the American Physical Society, Washington, D.C.

Coffman, W.E., Sex differences in responses to items in an aptitude test, Eighteenth Yearbook, National Council on Measurement in Education, 1961, 117 -24.

Corah, N.L., "Differentiation of children and their parents," Journal of Personality, 1965, 33, 300 - 308.

DeFries, J.C., Ashton, G.C., Johnson, R.C., Kuse, A.R., McClearn, G.E., Mi, M.P., Rashad, M.N., Vandenberg, S.G., and Wilson, J.R., Parent-offspring resemblance for specific cognitive abilities in two ethnic groups, Nature, 1976, 261, 131 - 33.

Denmark, Florence, What Sigmund Freud Didn't Know About Women and Men, Convocation Address, St. Olaf College, 1977.

Donlon, Thomas F., Content factors in sex differences on test questions, Research Memorandum 73-28. Princeton, N.J., Educational Testing Service, 1973.

Dreger, Ralph Mason and Aiken, Lewis R. Jr., The identification of number anxiety in a college population. The Journal of Educational Psychology, 1957, 48, 344 - 51.

Droege, R., Sex differences in aptitude motivation during high school, Journal of Counseling Psychology, 1967, 14, 407 - 11.

Dwyer, Carol, The role of tests and their construction in producing apparent sex-related differences in Andrisan-Wittig, M. and Peterson, A. (Eds), Sex-Related Differences in Cognitive Functioning, New York: Academic Press, 1979.

Elmore, Patricia B. and Vasu, Ellen S., Relationship between selected variables and statistics achievement: Building a theoretical model, Journal of Educational Psychology, 1980, August, Vol. 72,(4), 457 - 67.

Entwisle, Doris R. and Greenberger, Ellen, Adolescents' views of women's work-role, American Journal of Orthopsychiatry, 1972, 42, (4), 648 -56.

Erlick, Arline C., and LeBold, William K., Factors influencing the science career plans of high school students. Measurement and Research Center, Purdue University, West Lafayette, Indiana, June, 1975, 26pp.

Ernest, John, Mathematics and Sex, Santa Barbara: University of California, April, 1976.

Farley, Sister Mary de Chantal, A study of the mathematical interests, attitudes and achievement of tenth and eleventh grade students (Doctoral dissertation, University of Michigan, 1968). Dissertation Abstracts International, 1969, 29, 3039A. (University Microfilms No. 69-2312).

Fox, Lynn H., Career Interests and Mathematical Acceleration for Girls. Paper presented at the 1975 annual meeting of the American Psychological Association, Chicago, Illinois, August, 1975.

Fox, Lynn H., Gifted Girls: Scientists and Mathematicians of the Future. Paper presented at the 1976 annual meeting of the American Psychological Association of Gifted Children, Kansas City, Missouri, October, 1976.

Fox, Lynn H., Sex differences in Mathematical precocity: Bridging the gap. In Daniel P. Keating (Ed.), Intellectual Talent: Research and Development. Baltimore, Maryland: The Johns Hopkins University Press, 1976, 183 - 214.

Fox, Lynn H. and Denham, Susanne A., Values and Career Interests of Mathematically and Scientifically Precocious Youth. In J.C. Stanley, D.P. Keating and L.H. Fox (Eds.) Mathematical Talent: Discovery, Description and Development. Baltimore, Maryland: The Johns Hopkins University Press, 1974, 140 - 75.

Friedersdorf, Nancy Wheeler, A comparative study of counselor attitude toward the further educational and vocational plans of high school girls (Doctoral dissertation, Purdue University, 1969). Dissertation Abstracts International, 1970, 30, 422A (University Microfilm No. 70-03887.

Garai, J.E., Scheinfeld, A., Sex differences in mental and behavioral traits, Genetic Psychology Monographs, 1968, 77, 169 - 299.

Guilford, J.P., Fundamental Statistics in Psychology and Education. Fourth Education, (New York: 1965) McGraw-Hill.

Gulliksen, Harold, Theory of Mental Tests, (New York: 1950), John Wiley and Sons, Inc.

Harding, Jan, Sex differentiation and schooling, Education in Science, N89P27, Sept., 1980.

Hartlage, L.C., Sex-linked inheritance of spatial ability, Perceptual and Motor Skills, 1970, 73, 311 - 349.

Harway, Michelle, Astin, Helen, Suhr, Jeanne, and Whitely, John, Sex discrimination in guidance and counseling. National Center for Educational Statistics, Educational Division, U.S. Department of Health, Education and Welfare, February, 1972.

Haven, Elizabeth W., Factors associated with the selection of advanced academic mathematics courses by girls in high school (Doctoral dissertation, University of Pennsylvania, 1971, Dissertation Abstracts International, 1971, 32, 1747A (University Microfilms No. 71-26027)

Helson, Ravenna, Women mathematicians and the creative personality. Journal of Consulting and Clinical Psychology, 1971, 36, (2), 210 - 20.

Hoffman, Lois and Maier, Norman, Social factors influencing problem sol ing in women, Journal of Personality and Social Psychology, 1974, 29, 80 - 85.

Hilton, L.R. and Bergland, G.W., Sex differences in mathematics achievement -- a longitudinal study, Educational Testing Service, Research Bulletin Unpublished, 1971.

Hier, Daniel B. and Crowley, William F., Androgens and the development of spatial reasoning, New England Journal of Medicine, June, 1982.

Iglitzen, Lynne B., A child's eye view of sex roles. In Sex Role Stereotyping in the Schools, Washington, D.C., National Education Association, 1973, 23 - 30.

Jacobs, J.E., A comparison of the relationships between the level of acceptance of sex-role stereotyping and achievement and attitudes toward math of seventh graders and leventh graders ina suburban metropolitan New York community, Dissertation Abstracts International, 1973, 34, 3028A, (University Microfilms No. 73-28640)

Jay, Winifred Tom, Sex stereotyping in selected mathematics textbooks for grade two, four and six (Doctoral Dissertation, University of Oregon, 1973). Dissertation Abstracts International, 1973, 34, 3028A, (University Microfilms No. 73-28640).

Kaczala, Caroline and others, Developmental shifts in expectancies and attributions for performance in mathematics, RIE, May, 1980.

Kagan, Jerome, Spatial reasoning: Biology or environment, New England Journal of Medicine, June, 1982

Keeves, J.P., Differences between the sexes in math and science courses, International Review of Education, 1973, 11, 88 - 91

King, W.H., An experimental investigation into the relative merits of listening and reading comprehension for boys and girls of primary school age. British Journal of Psychology, 1959, 29, 42 - 49.

Levine, Maita, Identification of reasons why qualfied women do not pursue mathematical careers. Report to the National Science Foundation, August, 1976.

Levy, J., Lateral specialization of the human brain: Behavioral manifestations and possible evolutionary basis, In J.A. Kriger (Ed.) The Biology of Behavior, Corvallis, Oregon: Oregon University Press

Loehlin, J., Sharan, S., and Jacoby, R., In pursuit of the spatial gene: A family study, Behavior Genetics,January, 1978, 8, (1), 27 - 41.

Looft, William R., Sex differences in the expression of vocational aspirations by elementary school children, Developmental Psychology, 1971, 5, (2), 366.

Leinhardt, G., Seewald, Andrea, Engel, F., Learning what's taught: Sex differences in instruction, Journal of Educational Psychology, 1979, August, 71, (4), 432 - 439.

Luchins, Edith H., Women mathematicians: A Contemporary Appraisal, Paper presented at the 1976 meeting of the American Association for the Advancement of Science, Boston, February, 1976.

Maccoby, Eleanor E. and Jacklin, Carol N., The Psychology of Sex Differences, Stanford, California, The Stanford University Press, 1974.

McClure, W.C., A Multivariate inventory of attitudes toward selected components of elementary school mathematics (Doctoral dissertation), University of Virginia, 1979, Dissertation Abstracts International, 1971, 31, 5941A - 5942A.

Merkel, C., Sex differentiated attitudes toward math and sex-differentiated achievement in math on the ninth grade level in eight schools in New Jersey, (Doctoral dissertation) State University of New Jersey, 1974, Dissertation Abstracts International, 1974, 35, 3300A (University Microfilms No. 74-27331).

Milton, G.A., The effects of sex-role identification upon problem solving and skills, Journal of Abnormal and Social Psychology, 1957, 55, 208 - 213.

Milton, G.A., Sex differences in problem solving as a function of role appropriateness in problem content, Psychological Reports, 1959, 5, 705 - 8.

Mullis, Ina V.S., Educational achievement and sex discrimination, National Assessment of Educational Progress, 1975, 15pp.

Murphy, Roger J.L., Sex differences in examination performance: do these reflect differences in ability or sex-role stereotypes? Paper presented at the International Conference on Sex-Role Stereotyping, Cardiff, Wales.

Nevins, M., Sex differences in participation rates in math and science at Irish schools and universities, International Review of Education, 1973, 11, 88 - 91.

Ory, John C. and Helfrish, Linda, M., A study of individual characteristics and career aspirations, Paper presented at the annual meeting of the American Educational Research Association, San Francisco, California, April, 1976.

Perucci, Carolyn Cummings, Minority status and the pursuit of professional careers: Women in science and engineering. Social Forces, 1970, 19, (2), 245 - 259.

Petersen, A.C., "The relationship of androgenicity in males and females to spatial ability and fluent production," Unpublished Ph.D. dissertation, Department of Education, University of Chicago, 1973.

Plank, Emma N. and Plank, Robert, Emotional components in arithmetical learning as seen through autobiographies. Psychoanalytic Studies of the Child, 1954, 9, 274 - 96.

Poffenberger, Thomas M. and Norton, Donald A., Factors determining attitudes toward arithmetic and mathematics, The Arithmetic Teacher, April, 1956, 113 - 116.

Poffenberger, Thomas M. and Norton, Donald A., Factors in the formation of attitudes toward mathematics, Journal of Educational Research, 1959, 52, (5), 171 - 176.

Prediger, Dale J., McLure, Gail T. and Noeth, Richard, Promoting the exploration of personally relevant career options in science and technology, Report to the National Science Foundation, October, 1976.

Roberts, Fannie, Relationships in respect to attitudes toward mathematics, degree of authoritarianism, vocational interests; sex differences and scholastic achievement of college juniors, Dissertation Abstracts International, N31 (05), SecA, P2134.

Rogers, Margaret Anne, A different look at word problems, Mathematics Teacher, 1975, 68, (4), 258 - 88.

Rossi, Alice S., Women in science: Why so few? Science, 1965(a), 148 (3674): 1196 - 1202. Also in Toward a Sociology of Women, Lexington, Massachusetts: Xerox College Publishing, 1972, 141 - 53.

Rossi, Alice S., Barriers to the career choice of engineering, medicine or science among American women. In J. Mattfeld and C. Van Aken (Eds.), Women and the Scientific Professions, Cambridge, Massachusetts: M.I.T. Press, 1965 (b), 51 - 127.

Schlossberg, Nancy K. and Goodman, J., A woman's place: Children's sex stereotyping of occupations. Vocational Guidance Quarterly, 1972, 20, (4), 266 - 70.

Schlossberg, Nancy K. and Pietrofesa, J.J., Perspectives on counselors bias: Implications for counselor education, The Counseling Psychologist, 1974, 4, (1), 44 - 54.

Sells, Lucy W., The mathematics filter and the education of women and minorities, Paper presented at the annual meeting of the American Association for the Advancement of Science, Boston, Massachusetts, February, 1976.

Sherman, Julia, "Effects of Biological Factors on Sex-Related Differences on Mathematics Achievement" Women and Mathematics: Research Perspectives for Change (Washington D.C.: 1977) National Institute of Education.

Sherman, Julia, Predicting mathematics performance in high school girls and boys, grades 8-11, Journal of Educational Psychology, 1979, April, 71, (2), 242 - 49.

Sherman, Julia, Mathematics, spatial visualization and related factors: Changes in girls and boys, grades 8- 11, Journal of Educational Psychology, 1980, August, Volume 72, (4), 476 - 82.

Sherman, Julia A. and Fennema, Elizabeth, The study of mathematics by high school girls and boys: Related variables, Journal of Educational Research, 1976.

Sherman, Julia A. and Fennema, Elizabeth, Distribution of spatial visualization and mathematical problem solving scores: A test of Stafford's X-linked hypotheses, Psychology of Women Quarterly, 1978 Winter Volume, 3(2), 157 - 67.

Simpson, C.J., The effect of laboratory instruction on the achievement and attitudes of slow learners in math (Doctoral dissertation, LeHigh University, 1973), Dissertation Abstracts International, 1974, 34, 6959A-6960A, Microfilm No..74-11, 357.

Stafford, R.C., Sex differences in spatial visualization as evidence of sex-linked inheritance, Perceptual and Motor Skills, 1961, 13, 428.

Sternglanz, Sarah H. and Serbin, Lisa A., Sex role stereotyping as evidence of sex-linked inheritance, Developmental Psychology, 1974, 10, (5), 710 -715.

Strassberg-Rosenberg, Barbara and Donlon, Thomas F., Content influences on sex differences in performance on aptitude tests. Paper presented at the m eting of the National Council of Measurement in Education, 1975.