76-11,629

McCARTHY, Karen Ann, 1942-
SEX BIAS IN TESTS OF MATHEMATICAL APTITUDE.

The City University of New York
Ph.D., 1976
Education, psychology

Xerox University Microfilms, Ann Arbor, Michigan 48106

SEX BIAS IN TESTS OF MATHEMATICAL APTITUDE

by

KAREN McCARTHY

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

1975

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

12/19/75
date

Alan L. Gross
Chairman of Examining Committee

12/19/75
date

Executive Officer

Alan L. Gross

Patricia M. Kay

Max Weiner

Supervisory Committee

The City University of New York

ABSTRACT

SEX BIAS IN TESTS OF MATHEMATICAL APTITUDE

by

Karen McCarthy

Advisor: Professor Alan L. Gross

The primary objective of this research was to discover if the well documented differences in performance between males and females on tests of mathematical ability reflect _true_ differences in ability or stem, at least in part, from some artifact of the tests themselves. It was hypothesized that "bias" within the measures might stem from two possible sources: First, the subject matter content of the items, and second, the item tryout and selection procedures used in the construction of the tests.

The subjects of the study were 383 high school students in grades 10-12, enrolled in academic (i.e., college bound) mathematics courses. A 120 item pool of mathematics problem solving items was prepared, which varied from traditional item pools in that one-third of the items had content of familiarity and interest to females (e.g., cooking, sewing, etc.), one-third had content familiar to males (e.g., sports, building, etc.), and one-third were neutral in content. A sample of 50 students judged the extent of the differential familiarity of the item content to the sexes by rating it (i.e., the content) on a 1-5 Likert scale where 1 indicated

content of great familiarity to males, and 5, content of great familiarity to females. Mean ratings were computed, and the 78 items were selected for the final test which the students had seen most definitively as male, female, or neutral in content. (The balanced numbers of male-female-neutral were maintained.) This 78 item mathematics problem solving test was administered to the remaining 333 students.

Item analyses were conducted on the resulting data in three ways. Data for the total group (N=333) was analyzed and difficulty levels and point-biserial correlations for each item computed. The same computations were done on female data alone (n=148), and on male data alone (n=185). Using the traditional, commonly used approach to test construction, three tests were developed of 26 items apiece. The first test, known as the Total Group Test, was made up of the 26 items with the highest point-biserial correlations resulting from analysis of the data for the total group of subjects. It was found that among these items were 10 male, 4 female, and 12 neutral in content. The second test, known as the Male Group Test, was made up of the 26 items with the highest point-biserial correlations resulting from analysis of male data. The items were 11 male, 6 female and 9 neutral. The third test, the Female Group Test, was made up of the 26 items with the highest point-biserial correlations resulting from analysis of female data. For this third test the items were 2 male, 14 female, and 10 neutral.

In order to judge the similarity of these three tests, a measure of the extent of item "overlap" was computed. A comparison of the Total Group with the Male Group Test showed that 20 of the 26 items selected for the two tests were the same. That is, the tests had a 77% item correspondence. A comparison of the Total Group wih the Female Group Test resulted in a 15 out of 26, or 58% item correspondence, while the Male Group-Female Group Test resulted in a 13 out of 26, or 50% item correspondence. These outcomes indicated that the Total Group and Male Group Tests were essentially the same, while the Female Group Test varied substantially from both.

Subjects' raw scores were computed on all three measures and analyses of variance within sex subgroups were computed, followed by post hoc tests of differences. The male group means of 17.90 on the Total Group Test, 16.02 on the Male Group Test, and 13.98 on the Female Group Test were all found to be significantly different. Female students' means were 14.36 on the Total Group Test, 12.93 on the Male Group Test, and 15.61 on the Female Group Test. Among these, as among the males, all scores were found to be significantly different.

Analyses of Covariance between groups were computed where attitude toward mathematics, as measured by the Mathematics Attitude Scale (Aiken and Dreger, 1961), and Verbal Ability, as measured by the Differential Aptitude Test, were covaried alone and then jointly. These controls over attitude and verbal ability, however, did not change the significance of differences between males and females in performance. Males scored significantly higher than females on the Total Group and the Male Group Tests, while females scored significantly higher on the Female Group Test.

Three conclusions can be drawn from the research outcomes. First, when "best" items are selected from an item pool separately for sex subgroups, these items vary from those selected for the total group. Second, the variation in items is largely explained by the differential familiarity of the item content to the sexes. And third, sex subgroup performance on mathematics problem solving tests is higher when the test contains a high number of items with content familiar to that sex subgroup, and a limited number of items with content unfamiliar to the sex subgroup.

Further research is needed to investigate the developmental aspect of this phenomena. It would be of interest to discover if the same outcomes would eventuate among lower grade students where sex differences in mathematics are less pronounced and consistent, and whether the phenomena is progressive (i.e., grows more obvious as students get older). In addition, the question of the predictive validity of the existing "biased" measures need investigation, as does the validity of sex subgroup tests such as the ones here developed.

ACKNOWLEDGEMENTS

The author wishes to express her deepest thanks to Dr. Alan L. Gross for his unceasing encouragement, guidance, concern, and involvement throughout the past four years.

I am also grateful for the help and guidance provided by the other members of my dissertation research committee, Dr. Max Weiner, and Dr. Patricia Kay. Deserving special thanks is Dr. Carol Tittle, whose initial encouragement and advice provided the motivation for this research.

My further appreciation goes to the mathematics teachers and students at White Plains High School, without whose assistance and cooperation this study would not have been possible.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF TABLES
Continued

I INTRODUCTION

An area of particular interest in educational research today is the cause of the well documented difference in performance between males and females on tests of mathematical ability (i.e., the fact that males consistently score higher than females). A question frequently posed is: Do these differences reflect actual differences in mathematical ability, or can they be explained, at least in part, by some artifact of the tests themselves? Two possible artifacts which could contribute to differences in male-female performance are first, lack of familiarity with the subject matter of the items, and second, the item tryout and selection procedures used in the construction of the test. These two possibilities constitute the topic under investigation in this study.

If, in fact, the tests themselves influence the differential performance of the sexes, they can be said to be biased in some respect. Considerable research has been done in the area of test bias, much of which has dealt with bias in use for selection purposes (Cleary, 1968; Darlington, 1971; Thorndike, 1971.) In this research the bias under investigation was that which is reflected in inaccurate selection, or in other words that which results from the unfair <u>use</u> of the total test score rather than from the content of the test itself. Unlike these studies, the present research is concerned with the bias implicit in a test's items which might affect a subject's total test score, regardless of the intended use of that score. This type of bias may stem from two possible sources. The first of these is the subject matter content of the items. If one of the sex subgroups is less familiar with that content than the other, it is possible that this constitutes a source of bias against that subgroup. The second possible source of bias is the item tryout and selection procedure used in the construction of the test. The use of the point-biserial correlation as the index for selection of "best" items may become questionable in cases where a subgroup of the tryout population has scores which allow it to be clearly distinguished from another subgroup.

The question of the sex role appropriateness of subject matter content in problem solving items was dealt with by Milton (1958). Investigating differences in problem solving ability between males and females, Milton discovered that by altering the content characteristics of problems to make them less appropriate to the masculine sex role, sex differences in problem solving skill were reduced.

Similar results were found by Quirk and Medley (1972) in another context. Studying the relationship between race and subject matter content on the National Teachers Examination, these authors found that alteration of subject matter content to include items related to black culture caused significant improvement in the scores of members of this subgroup.

In still another study, Leder (1974) found that sex-determined leisure time activities can be translated to mathematics problem solving settings, and appropriately worded problems have a sex-determined differential meaningfulness to boys and girls which effects their performance.

In light of these findings, it seems apparent that if mathematics problem solving tasks are heavily weighted in item content familiarity toward one sex, then the test may be biased in favor of that subgroup.

The second possible source of bias, that resulting from item tryout and selection techniques, has been studied by Green (1972). Commonly, items are selected based on item tryouts with mixed sex groups. The typical procedure in selecting items is described by this author.

> "The first step is to develop a pool of items meeting various specifications as to form and content. Next these items are given to a sample of individuals. Various item statistics, such as point biserial correlations (item vs. total score), are calculated and the "best" items are then chosen; best is customarily characterized first and foremost by a high relationship of the item to the total score...Therefore, the items which discriminate best (i.e., show the highest relationship to total score) are the ones usually chosen (p.1)."

Investigating the existence of racial bias in achievement tests, Green looked for item by group interaction by selecting items separately for each of the racial groups from whom data had been collected on a total item pool. His results indicated that the use of tryout samples made up of population subgroups resulted in the selection of somewhat different sets of items for those subgroups. These findings might well be generalizable to the situation where mathematics problem solving items are selected for sex subgroups of the population, rather than for the total mixed sex groups commonly used in item tryouts.

The present study looks at the possible effects of these two potential sources of test bias on male-female performance on tests of mathematics problem solving ability. Based on the confirmed, existing differences in performance between males and females on these tasks, it is hypothesized that these differences in performance can be partially explained in several ways. Specifically:

(1) If "best"* items are selected from an item pool separately for sex subgroups, these items will vary from those selected as "best" for the total group.

(2) The variation in items will be partially explained by the differential familiarity of item content to the subgroups.

(3) Subgroup performance on "best" items selected using that subgroup as the tryout sample will be significantly higher than their performance on items selected for the other subgroup, or for the total group.

*Where "best" is defined as those items having the highest point-biserial correlations.

II REVIEW OF THE LITERATURE

The question of item content bias has been reviewed by a number of authors. Of particular interest is work with the Scholastic Aptitude Test (SAT) of the College Entrance Examination Board. This test has produced consistent differences in performance between males and females on its verbal and mathematics sections. Fremer and Chandler (1971) point out that differences between the sexes in performance were first noted at the test's earliest administration. In 1926, total scores were reported and females had a mean of 513 as opposed to 494 for males. In reviewing the individual items, however, it was discovered that males had scored higher then females in arithmetic problems and in number series completion.

Donlan (1971) reported on the 1930 testing where, for the first time, scores were reported separately for verbal and mathematics aptitudes. At that time females surpassed males in scores on the verbal portion of the test by 31 points. Males, on the other hand, surpassed females on the mathematical portion by 26 points. For a number of years this pattern continued. Coffman (1961) investigated the verbal differences and reported that women scored higher on items related to people, feelings, emotions, etc. Men succeeded more often on items related to things, mechanics, and science. Coffman concluded that, in general, verbal aptitude tests have a general content and since "things" tend to be more specific, verbal tests are more often people oriented, thereby giving females the advantage. Donlan (1971) confirmed Coffman's findings and further investigated item types. The SAT-V has four item types; antonyms and analogies where males and females have done equally well, and reading comprehension and synonyms where females tended to surpass males. In recent years, however, what had become the traditional superiority of females on the SAT-V has dwindled to insignificance. The 1966 and 1967 tests resulted in mean scores which indicated that the passage of time has, indeed, seen the verbal differences disappear. In 1966 males averaged 390 on the Verbal portion with females averaging 393. The 1967 Verbal results showed males with a mean of 463 and females with 464. Fremer and Chandler (1971) note that changes have been made in the content of the SAT-V over the years, but no content changes have been made that would account for the difference between recent findings and those characteristic of earlier years. Although sex differences in performance on the SAT-V no longer exist, the differences first evidenced on the SAT-M have grown as time has passed. In 1966, the mean score on the SAT-M for males was 422 while for females it was 382. In 1967 the male mean was 510, and the female 466. Thus, unlike the SAT-V results, differences between the sexes on the SAT-M have shown continued growth.

The SAT-M consists of 60 items in which there are only two formally distinct item types: general mathematics problem solving items common in form to many other tests, and data sufficiency items. Through the years there has been little change in item content. The test has moved away from the curriculum oriented type of items to items that depend on logical reasoning and on perception of mathematical relationships, but the content material in which the problems are imbedded has not been changed.

Donlan and Angoff (1971) discuss the fact that the setting of a mathematics item may be familiar or unfamiliar:

> "A setting is said to be familiar if it is similar to a setting which a typical candidate is likely to have encountered several times previously, either in actual experience or in a classroom exercise. The test for familiarity is the question, 'Is the candidate likely to accept the setting of this without puzzlement?' If the setting is familiar, the well-prepared candidate can proceed immediately to the solution of the problem without asking himself what the setting means (p.24)."

The question arises as to what happens when the well-prepared candidate is <u>not</u> familiar with the setting? Are females at a disadvantage on the SAT-M because the content (i.e., setting) of the test items is more familiar to males? The work of Donlan (1971) suggests that females are, indeed, at some disadvantage. Investigating the item difficulties of the SAT-M items, Donlan discovered that only two of the 60 items he reviewed favored females. In looking at the content of those two items he noted that one dealt with shirts being taken to a laundry, while the other was an algebra item with no content factor. Of the 60 items, 17 were referrents to "things", and no item was found to have a female agent... agents were "a boy", "John", etc. Donlan says that most of the items were everyday, neutral in content, but undisputably, "the overall effect is male (p.9)." Further, in a subgrouping of items, Donlan discovered that males missed ten percent fewer subject matter items than females, but only four percent fewer algebra items. This implies that the male advantage would diminish a great deal if only algebra (i.e., non-content) items were used.

A great many studies, not dealing directly with item content bias, but rather, confirming the existence of sex differences in mathematics performance have been conducted. Perhaps the most ambitious of these is the study by Husen (Ed.,1967). This author's work is a compilation of statistics and discussions of mathematics achievement internationally. Among the study's findings is the fact that "Sex was related to mathematics achievement in almost all countries, the boys scoring higher than the girls at all levels... (p.39)." The authors take the view that educational differences between the sexes are determined more by social and cultural differences than by innate differences in mental ability.

In an extensive review of the literature on mathematics learning and the sexes, Fennema (1973) cites numerous studies showing no significant differences between boys and girls in the pre-school years. In the early elementary grades sex differences begin to appear, but are inconsistent at best. Hervey (1966) found that boys were better able to solve problems presented verbally, while Lesser, Fifer and Clark (1965) found boys superior on space scales, and Wozencraft (1963) found girls superior in arithmetic reasoning. Heimgartner (1968) and Madden (1966) observed sex differences in mathematics ability as early as kindergarten.

As children enter adolescence the results of studies of mathematics achievement become less inconsistent. Girls are found to score higher on computational tasks and arithmetic fundamentals at this age, but boys appear to lead in reasoning ability (Unkel, 1966; Jarvis, 1964). Very (1967) states that male superiority in mathematics is due to greater exposure to the subject in the later years in school. Fitzsimmons (1969) points out that more boys than girls leave school at a younger age, and many of these are lower ability students, making the high school population of boys more homogeneous with respect to over all ability than that of girls. This may, in part, explain why boys, in the high school years, who choose mathematics courses, do better.*

*There is evidence (Wrigley, 1958; Shine, 1961) that there is a high positive relationship between mathematics ability and general intelligence.

Aiken (1971) presents evidence that individual sex differences in mathematics ability increase steadily at successive ages or grade levels.

It becomes obvious in reviewing this literature that findings and opinions about sex differences in mathematics are, at best, confusing and inconsistent. The one fairly well undisputed fact is that by the late high school years boys are surpassing girls consistently on tests where arithmetic reasoning plays a strong part.

Looking at a factor analytic study of mathematics, Very (1967) and Coleman (1956) found that there are three components to mathematics:

1) Arithmetic knowledge-number of courses, degree of achievement or sophistication

2) Numerical ability-computational operations

3) Quantitative reasoning-verbal problem solving ability

A series of studies by Sweeney (1953) revealed that sex differences in mathematics were not related to intelligence, spatial ability or knowledge, or numerical ability, but primarily to the quantitative reasoning factor. Boys appear more able to restructure problems presented,,,a skill needed for the successful solution of problems. It would appear, therefore, that from adolescence, mathematics sex differences are due primarily to differences in problem solving ability. A variety of research studies have been conducted into problem solving ability and these will help to shed more light on our question.

Berry (1958) was one of the earliest to conclude that attitudes toward problem solving were more favorable among men, and that these attitudes were clearly correlated with problem solving ability. This relationship between attitude and problem solving ability has been firmly established by a number of studies. Brown and Abel (1965) demonstrated that the correlation between pupil attitude toward a subject and achievement in that subject is higher for arithmetic than for spelling, reading, or language. Aiken and Dreger (1961) had shown that scores on a mathematics attitude scale contributed significantly to the prediction of final mathematics grades for a group of 67 female subjects, but the results were not significant for male subjects. Aiken (1963) showed that for 160 college women the correlation between mathematics attitude scale score and SAT-V was not significant, between the mathematics attitude scale score and SAT-M the correlation was significant. Reese (1961) has found that, in general, measures of attitude and anxiety are better predictors of female achievement than male. Aiken (1972) has found that attitude toward mathematics is inversely related to language arts, social studies and other "verbal" pursuits. In addition to his findings concerning the relationship between attitude and problem solving ability, Berry (1958) also found that women's scores on the Terman-Miles Masculinity-Femininity (M-F) scale were significantly correlated with this ability (i.e., a woman with more masculine interests has a more favorable attitude toward problem solving). Earlier, Carey (1955) had found that women, as a group, do dislike problem solving, but when exposed to discussions to improve their attitudes their problem solving skills improved.

This leads us to the work of Milton (1958), mentioned previously, which is among the most significant in terms of our question. Milton first hypothesized that problem solving ability is positively correlated with degree of indentification with the masculine role. He further hypothesized that this characteristic occurs within and across sex, and that an adjustment for between subject variance of sex role identification makes problem solving differences disappear. Milton administered a 20 problem test, the Terman-Miles M-F scale (for sex role identification), and the Otis Quick Scoring Intelligence Test, to 37 female and 42 male high school students. Using an analysis of covariance technique to control for sex role identification, differences in problem solving ability did indeed disappear. Milton then became interested in the role appropriateness of problem content. He had found that most measures of problem solving relied heavily on questions of greater familiarity to subjects with masculine sex role identifications. Consequently, he hypothesized that when the characteristics of problems are altered to make them less appropriate to the masculine sex role the sex differences in problem solving skill will be reduced. It was first necessary to determine the nature of masculine and feminine role content. Its existence had long been established. As early as 1936, Terman and Miles had reported consistent sex differences in interests. This had been further confirmed by Hartley and Klein (1957). Accordingly, Milton had 60 Stanford undergraduates (30 male, 30 female) report the kinds of problem solving they encountered each day. The problems were then categorized by sex by two independent judges and sex differences in content were noted. The author then developed two sets of problems (40 items), one set of which were conventional to problem solving literature, the other set in a form more appropriate to the female sex role. These feminine items were constructed to be parallel to the masculine items (i.e., the task was the same, but the content was altered to be feminine in nature). The items were randomly presented in a 20 item test (10 masculine, 10 feminine), but no test had the same task in both sex role forms. The Terman-Miles M-F scale was administered to 25 male and 25 female undergraduates as was the problem solving test. The results of the experiment confirmed the hypothesis. Males solved, on the average, more problems than did females, but the difference between males and females is reduced when the problems are framed in a content appropriate to the feminine role. The results suggested that to the extent that problems can be made more appropriate to the feminine role, or less appropriate to the masculine role, the sex differences in problem solving achievement are diminished. Milton concluded

> "It is not entirely that men have a better developed 'general reasoning' capacity, or that they have learned more skill, but appar-

> ently that they are responding in part merely to the stimulus properties, which in the case of problem solving research has been predominantly appropriate to the male role (p.15)."

He goes on to say

> "...the findings also suggest an important implication concerning the improvement and measurement of the problem solving skills in women. In order to achieve maximum results in either case, it seems advisable to frame the problem solving tasks in content appropriate to the sex role. This may, in many ways be more efficient than trying to convince women that problem solving of the conventional sort <u>is</u> role appropriate (p.15)."

A later replication confirmed the results of this study.

In another experiment Milton (1958) explored the strategies used to solve problems by the two sexes. He lists three apparent strategies for problem solving: (a) Seeking help (b) Trial and error (c) Analysis or abstraction. He hypothesized that (a) males and females would differ in choice of strategy (b) an interaction between strategy and sex role appropriateness of content would be seen and (c) the relationship between choice of strategy and sex role identification within sex would be similar to that between sexes. Using 40 male and 40 female undergraduates, Milton administered a 16 item forced choice test. Each item had one problem with three alternative methods for solving it each corresponding to the three strategies. Subjects had to choose the alternative they would use to solve the problem and also indicate which they felt was the "best" alternative.

Among the 16 items, eight were male in content and eight were female. The findings indicated that men more often than women chose the analytic strategy to use, but women more often than men said the analytic was best. Most frequently women chose the trial and error method for use. These findings led Milton to the belief that the superior achievement of men in problem solving could be accounted for by their greater tendency to use an analytic strategy (which is a superior strategy).

Another explanation of sex differences in problem solving mathematical tasks is proferred by McMahan (1972). McMahan was interested in the causal attribution and expectations of success in regard to age and sex differences. His subjects were 325 students in grades 6, 10, and first year of college, approximately equally divided by sex. He presented each age group (separately) with a set of scrambled word problems and a set of arithmetic problems. The subjects were assured that they had learned how to do the types of arithmetic problems and that they would be familiar with all of the words in the scrambled word test. He also had no time limit set for either task. Subjects were asked, prior to testing, to predict how many they would get right. The results showed no differences in expectations of success on the scrambled word problems, however, females at every grade level underpredicted their performance on arithmetic items. The author interpreted this as evidence of the effect of sex stereotyping on the anticipation of success.

McMahan's conclusions are a further extension of the findings of Horner (1968). She had stated that females who find themselves in competitive situations are in conflict between obtaining the positive reward of success and suffering negative societal sanctions for having been "unfeminine" should success eventuate. This conflict is even greater when the competition is male and when the performance situation involves activities conceived of as traditionally more masculine than feminine. If we can assure from Milton's (and others) work that mathematics may be considered "traditionally more masculine than feminine" then females would face this conflict of need to succeed, need to fail in any mathematics activity. This need for achievement, fear of failure, fear of success conflict had also been postulated to account for differential levels of performance in achievement contexts by Atkinson and Feather (1966).

To summarize the research cited, it has been shown that findings on early sex differences in mathematical ability are inconsistent, but in general, the difference between the sexes in problem solving ability is firmly documented by the late high school years. In exploring the area of problem solving, researchers have found that it is traditionally thought of as a masculine activity, and in fact, problem solving tasks are most often heavily male in content. Females have been shown to believe they are less able to do well in mathematics (problem solving) and it has been hypothesized that they may do less well, in part, because of their fear of success on a masculine task (labeling them as unfeminine).

The relationship between attitude toward problem solving and performance has been clearly established, with women evidencing significantly poorer attitudes toward the activity than men. Two methods of improving women's performance have been suggested and shown to be successful. The first is based on the assumption that is impossible to change a woman's attitude toward the problem solving task, so instead the task must be changed. In this case, the problem solving items are made female in content (although the problem itself is unchanged) and women then score significantly higher. The second involves attempts to make women view the problem solving activity itself more favorably and discussion groups to this end did succeed in improving women's scores on the problem solving tasks.

In order to better understand the growth of differences between males and females in achievement, it is necessary to review the literature on the development of sex differences in children. This requires an understanding of the development of an individual's sex-role identification, a long and complicated process beginning in some researchers eyes, even before birth. Lewis (1972) has said that women even react in a sex appropriate fashion to the behavior of the fetus. If it is a particularly active fetus, moving and kicking with some strength, the woman tends to think of it as masculine. There are substantial differences of opinion about how sex differences arise, but no controversy about whether they exist. Lewis and other (Messick, in print: Kagan, 1972 etc.) believe that parents treat children differentially by sex from the moment of birth. He tells us that observation of parents' behavior towards their infants reveals that the types of parental attachment behavior directed toward the infant varies directly as a function of the child's sex. Moss (1967) and Yarrow, Rubenstein, and Pederson (1971) tell us that there is evidence that parents handle boys more roughly than girls. Newson and Newson (1968) noted that boys receive more physical punishment than girls and are often simply told "NO" while girls are presented with alternatives for behavior. A discussion by Lynn (1972) centers on the modeling of young children. He points out that in the early years children of both sexes spend the vast majority of their time with their mother. Girls in their early years may appropriately copy their mother's behavior...learning their behavior patterns rotely. Boys, on the other hand, are unable to do this. In learning the mother-identification lesson, girls have had little practice in ignoring irrelevant cues and isolating significant ones, whereas boys have had much practice in doing so while solving the masculine-role identification problem. This lesson, learned very early in life, is one of a boy's earliest encounters with the problem solving situation. Lynn believes that the differences in early parent treatment of the sexes and the early experiences of boys and girls cause them to develop different cognitive styles. The boy acquires a cognitive style that primarily involves (1) defining the goal, (2) restructuring the situation and (3) abstracting principles. In contrast, the girl acquires a cognitive style that primarily involves (1) a personal relationship and (2) lesson learning. This belief, that the environment of the child in the earliest years causes the sexes to develop different cognitive styles, is not universally accepted. Maccoby and Jacklin (1972) and Maccoby (1972) claim that the foundations of sex differences aren't laid until after the age of six. Maccoby claims that there is no evidence that the modeling process takes place at all. She is willing to state, however, that in the later years (after six), parents do begin to reinforce children of different sexes differentially.

Despite Maccoby's claims, there appears to be more evidence that differences do appear in the earliest years, and that these differences are, at least in part, mediated by differential treatment by parents and others in the young childs' environment.

In the school years, differences in treatment by sex become more pronounced and are better documented. In the last few years numerous studies of school materials have been done pointing out the possible patterning of the child by the overwhelming presence of sterotypes in school materials. Gunderson (1972) has documented the fact that reading difficulties in the early school years are far more prevalent among boys than girls. Ninety percent of referrals to reading clinics are males. One possible explanation is that adults consider reading a feminine activity (i.e., it doesn't entail physical movement). Carroll (1971) says that the "nearly universal" finding of early language ability in girls in this country is due to characteristic differences between the ways boys and girls are reared and socialized. Preston (1962) say that boys are superior in reading in Germany where there are more male teachers in the early grades and different cultural attitudes. Although early grade basal readers hold more interest for girls, they also perpetuate the stereotyped female role. Women seldom, if ever, work outside the home and men never work in it. Boys are given roles of leadership, girls are followers; boys are active, girls are passive etc. (Howe, 1971). In later school readers the emphasis is on male stories, male biographies and stereotypic roles. Among outstanding examples of children's literature, Sadker's (1973) study of Newberry Award books showed they were 3-1 about boys. The most common young girl role is the "copout"...that is, girls who were tom-boys, but reformed and became sweet and feminine to win the boy. Nilsen (1971) analyzed the Caldecott Award winners and found 10 had girls as leading characters while 24 had boys. Book lists recommended by the Child Study Association as "especially for girls" were mainly about love, dating and romance, or personal problems. Child, Potter, and Levine (1946) found that boys in readers attempted to solve problems (which girls often present them with, having been unable to solve them themselves). Malcolm (1971) found that in a study of school books (grades 1-8) the word boy or boys appeared 4,700 times versus 2,200 for girl or girls, and of the 20 given names appearing most frequently 13 were male and 7 female. The literature shows no study of mathematics texts specifically, but a study of standarized achievement tests (Tittle, McCarthy and Steckler, 1973) showed the same results as reported in school texts... overwhelming male references and stereotyping.

And what is the effect of this preponderance of stereotyping on young children? Dr. Edward Hall in <u>Silent Language</u> (1959) says children assimilate the content and values of their books as they learn to read without giving it conscious thought. Watson (1959), Hartley (1959), and Baumrind (1972) say that by the age of eight, there is 99% agreement among children of both sexes as to which sex does what job, what kind of person a boy or a girl should be and what their role limitations and expectations are. Sadker's (1973) study showed that girls in the upper high school grades envision themselves in four careers primarily: elementary school teacher, nurse, secretary, or mother. In another study cited by Sadker, boys were shown to be willing to persevere and face possible failure. In this study upper elementary school children were given four puzzles to solve and told they had a time limit. When the given time was up, those children who had not successfully completed all puzzles were told they had some extra time and could do whatever they wanted. Boys significantly more often continued work on the unfinished or incorrectly done puzzle while girls chose to re-do one of those already done correctly.

Dwyer (1972) in a study of sex role standards and sex role identification and their relationship to achievement found that children's sex role standards were positively and significantly related to reading and arithmetic while the sex of the subject and liking or disliking reading and arithmetic were not. This suggests that reading and arithmetic sex differences are more a function of children's perceiving reading as an activity appropriate to girls and arithmetic as an activity appropriate to boys, than of the students' biological sex, individual preference for the masculine or feminine sex role, or liking or disliking of reading or arithmetic. (The children's perception of subject matter areas of reading and arithmetic as masculine or feminine contributed significant variance to their achievement test scores in reading and mathematics.)

Certainly languages add to the problems. As Key (1970) points out, in English there is no singular equivalent of the neuter, plural "they." The French language presents even more ludicrous problems. The designation of nouns as masculine or feminine shows the words doctor and professor to be masculine. Consider the sentences:

Le docteur, elle est dans le salon.
↓ ↓
masculine she

or

Le professeur est enciente!
↓ ↓
Masculine pregnant!

In English, however, the work of Tittle, McCarthy and Steckler (1975) studying sex referenced words in school achievement tests, showed that generic words did not contribute significantly to the number of sex referenced words present in the tests.

A summary of the research in this area shows that the preponderance of evidence points to early sex differences, contributed to in part by differential parental treatment of boys and girls. Overwhelming evidence of school material stereotyping exists, not to speak of continued stereotyping of women by public figures, employers, the media etc. Empirical evidence has been documented showing that boys and girls succumb to the stereotyping by believing that boys and girls do and should fit the roles set for them. This extends to a belief, on their part, that reading is a feminine activity and arithmetic a masculine one.

In summation, it is possible to interpret the research on the development of sex differences in mathematics achievement as follows:

Male and female infants from the moment of birth are patterned (socialized) to fit certain explicit sex roles. The patterning begins with parents and others in the child's environment and continues (and is perhaps more emphatic) as the child enters school. The male child by the time he reaches school has already had extensive experience with problem solving, and exerting independence, and he has been taught that he is expected to achieve. The female child has developed a pattern of rote learning, dependence and submission. Because the early school years reward children for rote learning and docile, cooperative behavior, girls are, in general, "better" students. The materials presented to the child in the early school years teach the child further the appropriate role she/he is to play. Among other ideas being reinforced is one which states that reading is a feminine activity while mathematics is a male activity. Nevertheless, the roteness of early arithmetic tasks allows girls to excel in these computational activites. Meanwhile, boys, in all of their activities, because of their greater participation in active rather than passive roles, are having more opportunities to deal with problem solving while girls are continuing with imitation and follower activities. In the later school years the tone of textbooks shows males taking leadership roles, achieving intellectually and professionally, while women are stereotyped into the mother-nurse-teacher-secretary life of serving. By the last years of high school both girls and boys have, in good part, accepted their assigned roles. Girls believe they are unable to solve problems and that it is inappropriate for them to be "high achievers" particularly in activities where they are competing with boys. It is appropriate, however, for them to show themselves as able in verbal pursuits and so they continue to achieve equally well (to males) on verbal tasks. Quantitative tasks have, to them, become male and their negative attitude toward these tasks sorely affects their performance on these problem solving activities. (The effect of attitude on problem solving has been clearly documented.)

Minuchin (1972) has presented four goals for education to help reverse this trend:

(1) Stereotypes in school materials must be minimzed.

(2) Girls must be provided with broad exposure to experiences, ideas, and female models who vary from the sterotyped model.

(3) Girls must be educated in skills for choice, problem solving and evaluation.

(4) Girls must be encouraged in self-differentiation and self-knowledge.

Fulfillment of these goals in American schools is bound to take considerable time, and in the interim it is essential that we not limit women's options through the use of measures which may be unfair. Two possible sources of bias on mathematics tests are the documented "maleness" of the content of the items, and the item selection techniques employed in the test construction. It is possible, as suggested by Donlan (1971) that a balance in the sex-orientation of the content would cause females' scores to more accurately reflect their abilities. Further, it may be that, as suggested by Green (1972), the item selection techniques commonly used in test construction are also contributing to bias against females. These two potential sources of bias indicated a need for exploration.

III METHOD

Subjects

The subjects of this study were 383 high school students in grades 10 to 12. They represent the total enrollment in academic (i.e., college bound) mathematics classes at a large suburban consolidated high school. The experimental group was racially and ethnically mixed in the same proportions as the total school population and contained approximately equal numbers of males and females.

Procedure

The experimental procedure was divided into seven stages.

In Stage 1, an item pool of 120 mathematical problems, similar in format (i.e., multiple choice) and task content to those found on the SAT-M was constructed. The item pool varied in form from more traditional mathematics problem solving item pools, however, in the respect that it contained what could be thought of as three distinct "sets" of items. The first set contained 40 items with content intended to be more familiar to females than to males. That is, the subject matter language in which the problems were imbedded dealt with areas with which females would be likely to have had greater experience. The second set had 40 items with content of greater familiarity to males than females, and the third set had 40 items which were intended to be neutral in content. The mathematical tasks in each set were matched in terms of skills needed for solution, and difficulty. That is, each set of items contained approximately equal numbers of problems involving basic arithmetic functions, percents, decimals, fractions, simple algebraic functions, etc. In addition, the attempt was made to see that no one set of items was, on the average, of greater difficulty than any other set.

In Stage 2, a panel of six high school mathematics teachers reviewed the 120 items for verification of the following points;

(1) The three sets of items were of approximately equal difficulty.

(2) The three sets of items required similar skill development for solution, and

(3) All items were appropriate to the mathematics achievement levels of college bound high school students in grades 10 to 12.

In Stage 3, upon verification of the above points, the 120 item pool (See Appendix I) was submitted to two classes which had been selected randomly from the entire subject sample. These two classes contained a total of 50 students, 28 of whom were male and 22 female. These 50 students were asked to judge the items as to the differential familiarity of the content to male and female students of their age. The 120 items were randomly ordered and presented without the multiple choice answers. Subjects were asked to read the item with a view not toward solution of the problem, but rather, toward judging how familiar the item content was to males and females of their ages. Each item was followed by a 1-5 Likert Scale, where 1 signified content of great familiarity to males, and 5 suggested content heavily familiar to females. A score of 3 would indicate content of equal familiarity to males and females.

Mean judgement scores were computed for each item separately for the two sex groups. That is, the average Likert Scale rating indicating the male subjects' judgement as to the differential familiarity of the item content to males and females was computed, and the same computation was done based on female subjects' judgements. In addition, for the total group of 50 students (i.e., males and females pooled) averages were computed. Items receiving judgement means in the range of 1.0 to 2.0 were retained and designated as male in content. Those items with judgement means in the 4.0 to 5.0 range were designated as female, and those in the 2.5 to 3.5 range were designated as neutral. All items with means not within the indicated ranges were dropped from the item pool, as they were considered ambiguous with regard to their differential content familiarity to the sexes. Further, male subjects' judgement ratings were compared with female subjects' judgement ratings. In the case where male and female subjects differed in their judgement rating, the item was also eliminated. That is, if male subjects designated an item as female, for example, but female subjects designated the same item as neutral, the item was dropped from the item pool.

At the completion of this stage, 78 items remained in the item pool; 26 rated by the student judges as male in content, 26 as female in content, and 26 viewed as neutral in content. These items constituted the final test item pool. (It is to be noted that the procedure described in Stage 3 was developed in a Pilot Study which is described in Appendix II.)

Stage 4 of the study involved the testing of the remaining members of the subject group. Three hundred thirty-three students, 185 male, and 148 female, were administered the Mathematics Attitude Scale (Aiken and Dreger, 1961; See Appendix III) by their classroom teachers. Two weeks later these same students were administered the 78 item mathematics problem solving test.

The 78 items which had remained after the initial elimination of items, had been arranged in two 39 item tests (See Appendix IV). Items were selected randomly from the 26 male, 26 female, and 26 neutral items to form the two 39 item tests, each containing three sets of 13 items balanced by sex designation. The division of the test into two parts was done to facilitate administration. Tests were administered in the 50 minute class period by the mathematics teachers on two consecutive school days. Instructions and explanations to the students were standardized across the classes. Although administered in two parts, the analysis of results was conducted on the full 78 item pool. That is, tests were scored as though the highest possible raw score was 78.

Stage 5 involved a series of three item analyses conducted on the data resulting from the Stage 4 test administration. The item analysis included computation of difficulty levels (p) and point-biserial correlations (r_{pbis}) for each of the 78 items. In addition, the analysis computed the mean, standard deviation, standard error of measurement, and the Kuder-Richardson 20 reliability coefficient for the total test.

The first item analysis was done utilizing the data from the total 333 student subject population. Analyzing the data simultaneously for both males and females is the procedure most commonly used in test construction. After computation of the 78 point-biserial (i.e., item-total score) correlations, and associated difficulty levels, the 26 items with the highest r_{pbis} were identified and designated as Test 1, the Total Group Test of 26 "best" items.

It should be noted that the decision to construct tests of 26 items apiece was made because 26 constitutes one-third of the total item pool. This total item pool is made up of one-third male, one-third female, and one-third neutral items. In constructing tests separately for the two sex subgroups, where it is hypothesized that an interaction between subgroup membership and item sex content may exist, theoretically, for example, all 26 male items might be selected when male data is analyzed separately.

The second item analysis was conducted on data for male subjects (n=185) only. Again, 78 r_{pbis} were computed and the 26 items with the highest correlations were identified and designated as Test 2, the Male Group Test of 26 "best" items.

The third item analysis was conducted on data for female subjects (n=148) only. The 78 r_{pbis} were computed and the 26 items with the highest correlations were isolated and designated as Test 3, the Female Group Test.

Stage 6 of the procedure was intended to discover the extent of item "overlap" between tests. That is, to discover the extent of item commonality between the tests. In order to discover if the "best" items for the total group were the same as the "best" items for the female group, for example, a set of pairwise comparisons were made:

(1) Items chosen as "best" for the total group compared with items chosen as "best" for the male group (Test 1 vs. Test 2).

(2) Items chosen as "best" for the total group compared with items chosen as "best" for the female group (Test 1 vs. Test 3).

(3) Items chosen as "best" for the male group compared with items chosen as best for the female group (Test 2 vs. Test 3).

The number and percentage of items chosen as "best" for one member of a pair, but not for the other was recorded. This percentage can be thought of as representing the extent of interaction between the items and the group membership factor. Items selected for one member of a pair and not for the other can also be thought of as "biased", as they represent "best" items for one subgroup, but not for the other.

Also included in this stage of the analysis was a computation of the mean sex judgement rating for each test. It was of interest, for example, to discover if the test items selected using male group data only were more male in content than the test items selected using female group data. In addition, a simple count of items based on sex content designation was computed for each test.

Stage 7 of the analysis consisted of the computation of male and female performance (separately) on each of the constructed tests. Because all subjects had taken all items initially, it was possible to compute means and standard deviations for the male group and the female group on the three 26 item tests constructed under the conditions described in Stage 5. Stage 7 resulted in the following performance means:

(1) Mean performance score for all male subjects on Test 1, the Total Group Test.

(2) Mean performance score for all male subjects on Test 2, the Male Group Test.

(3) Mean performance score for all male subjects on Test 3, the Female Group Test.

(4) Mean performance score for all female subjects on Test 1, the Total Group Test.

(5) Mean performance score for all female subjects on Test 2, the male group test.

(6) Mean performance score for all female subjects on Test 3, the female group test.

After computation of the above means, it was of interest to discover if performance varied significantly between the three tests. That is, did females (or males) score significantly higher or lower on any one test? To this end, two analyses of variance (ANOVA) were computed. The first compared mean performance scores for males on the three tests, while the second compared the mean performance scores for females on the three tests. These ANOVA's were followed by post hoc tests to identify the specific means which varied significantly from one another.

Finally, male and female performance was compared on the three measures in three further ways. An analysis of covariance between male and female performance on the three measures was computed using Mathematics Attitude Scale score as the covariate. Earlier research indications that attitude toward mathematics is a significant factor in achievement (Berry, 1958; Aiken and Dreger, 1961; Brown and Abel, 1965; Aiken, 1963) made it imperative that this variable be removed from consideration when comparing male and female performance.

A second analysis of covariance between male and female performance on the three measures was computed using Verbal score on the Differential Aptitude Test as the covariate. This was done to control for sex differences in reading ability which might confound problem solving ability, as suggested by Spilman and Weiner (1972).

The third analysis of covariance was computed using both Mathematics Attitude Scale scores and Differential Aptitude Test Verbal scores as covariates.

IV RESULTS

The administration of the 120 item pool to a random sample of 50 of the subjects in Stage 3 of the study resulted in three sets of sex content judgement rating means: The mean sex judgement rating given the item by male subjects ($\bar{X}_m$), the same mean as given by female subjects ($\bar{X}_f$), and the mean for the total, pooled sex group ($\bar{X}_t$). These mean ratings and their associated standard deviations are shown in Table 1. In each case (i.e., for each item) it was noted whether agreement existed between the male and female judges. That is, if the male judges gave the item a mean sex judgement rating in the neutral range, it was noted whether females also gave the item a neutral rating. Disagreement between male and female subjects occurred on 18 of the 120 items (e.g., item #4, rated as ambiguous by males, but neutral by females). These 18 items were the first to be eliminated from the item pool. For the 102 items remaining, the ratings of the total, pooled sex group were used for decision-making. Fourteen of these 102 items had been rated as ambiguous, and these too were dropped from the item pool. Remaining after these eliminations were 36 neutral, 26 male, and 26 female items. In order to maintain the sex content balance of the item pool, it was decided to eliminate 10 of the neutral items. As a 3.00 rating would indicate maximum neutrality, it was decided that those items which

TABLE 1
Sex Content Judgement Ratings With Associated Standard Deviations for 120 Item Pool

	Male Group n=28		Female Group n=22		Total Group n=50		Designation
Item Number	$\bar{X}$m	SDm	$\bar{X}$f	SDf	$\bar{X}$t	SDt	
1	2.96	.18	3.00	.00	2.98	.14	Neutral
2	3.17	.66	3.00	.39	3.10	.54	Neutral
3	3.92	.67	3.72	.63	3.82	.69	Ambiguous
4	3.57	.69	3.50	.59	3.54	.64	Ambiguous
5	2.28	.71	2.54	.59	2.40	.67	Ambiguous
6	1.60	.62	1.50	.59	1.56	.61	Male
7	2.65	.44	2.95	.21	2.65	.36	Neutral
8	4.39	.62	4.54	.59	4.46	.61	Female
9	2.10	.73	2.31	.64	2.20	.69	Ambiguous
10	3.53	.63	3.90	.61	3.70	.64	Ambiguous
11	2.96	.42	3.00	.00	2.98	.31	Neutral
12	4.46	.57	4.45	.59	4.46	.57	Female
13	3.78	.73	3.81	.58	3.80	.67	Ambiguous
14	3.64	.78	3.59	.50	3.62	.66	Ambiguous
15	1.42	.57	1.63	.65	1.52	.61	Male
16	1.71	.71	1.68	.56	1.70	.64	Male
17	2.53	.83	2.90	.71	2.62	.77	Neutral
18	3.03	.70	2.81	.58	2.91	.67	Neutral
19	2.21	.73	2.31	.56	2.26	.66	Ambiguous
20	2.89	.78	2.90	.29	2.89	.62	Neutral
21	1.82	.66	1.81	.58	1.82	.62	Male
22	2.53	.79	3.00	1.11	2.54	.81	Neutral
23	4.32	.94	4.45	.50	4.40	.76	Female
24	3.18	.55	3.00	.30	3.08	.48	Neutral
25	4.39	.83	4.63	.58	4.50	.73	Female
26	1.71	.65	1.45	.67	1.60	.67	Male
27	4.10	.83	4.68	.47	4.36	.74	Female
28	2.85	.35	3.00	.00	2.92	.27	Neutral
29	3.71	.65	3.61	.66	3.68	.65	Ambiguous
30	2.57	.79	2.68	.47	2.77	.65	Neutral
31	3.10	.68	3.22	.52	3.16	.61	Neutral
32	4.10	.87	4.22	.61	4.16	.76	Female
33	3.86	.83	4.04	.72	3.96	.78	Ambiguous
34	4.07	.76	4.31	.89	4.18	.82	Female
35	4.14	.93	4.40	.73	4.26	.85	Female
36	2.10	.73	2.04	.72	2.08	.72	Ambiguous
37	2.25	.75	2.63	.58	2.42	.70	Ambiguous
38	4.39	.73	4.27	.63	4.34	.68	Female
39	2.53	.63	2.59	.79	2.56	.70	Neutral
40	1.88	.75	1.90	.81	1.90	.76	Male

TABLE 1
Continued

Item Number	Male Group n=28		Female Group n=22		Total Group n=50		Designation
	$\bar{X}m$	SDm	$\bar{X}f$	SDf	$\bar{X}t$	SDt	
41	4.25	.70	4.36	.72	4.30	.70	Female
42	2.92	.46	3.04	.21	2.98	.37	Neutral
43	1.82	.81	1.86	.63	1.81	.72	Male
44	1.54	.72	1.42	.59	1.50	.70	Male
45	2.50	.69	2.31	.64	2.42	.67	Ambiguous
46	2.14	.75	2.54	.59	2.32	.71	Ambiguous
47	4.25	.64	4.36	.58	4.30	.61	Female
48	2.53	.63	2.36	.65	2.46	.64	Ambiguous
49	2.92	.26	3.00	.00	2.96	.19	Neutral
50	3.67	.66	4.09	.53	3.84	.65	Ambiguous
51	1.85	.84	1.95	.57	1.90	.73	Male
52	1.70	.91	1.90	.68	1.82	.82	Male
53	2.10	.83	2.40	.50	2.24	.71	Ambiguous
54	1.96	.96	1.63	.72	1.82	.87	Male
55	4.46	.74	4.54	.67	4.53	.68	Female
56	2.92	.78	2.90	.29	2.90	.61	Neutral
57	1.67	.98	1.54	.67	1.62	.85	Male
58	1.74	.90	1.77	.61	1.78	.78	Male
59	4.03	.92	4.27	.55	4.14	.78	Female
60	4.07	.85	4.45	.67	4.24	.79	Female
61	3.21	.73	3.09	.29	3.16	.58	Neutral
62	4.07	.66	4.31	.56	4.18	.62	Female
63	3.92	.85	3.80	.81	3.94	.84	Ambiguous
64	4.07	.71	4.50	.59	4.26	.69	Female
65	4.21	1.03	4.31	.56	4.26	.85	Female
66	1.67	.77	1.81	.66	1.74	.72	Male
67	2.53	.57	2.77	.42	2.64	.52	Neutral
68	1.50	.63	1.63	.72	1.56	.67	Male
69	2.92	.37	3.04	.21	2.98	.31	Neutral
70	2.35	.62	2.40	.50	2.38	.56	Ambiguous
71	2.10	.62	2.09	.52	2.10	.58	Ambiguous
72	3.17	.47	3.09	.29	3.14	.40	Neutral
73	1.46	.57	1.27	.55	1.38	.56	Male
74	4.11	.80	4.68	.47	4.32	.85	Female
75	3.00	.27	3.04	.21	3.02	.24	Neutral
76	3.35	.73	2.86	.35	3.30	.64	Neutral
77	2.53	.50	2.81	.39	2.66	.47	Neutral
78	2.57	.63	2.72	.45	2.64	.56	Neutral
79	4.10	.56	4.09	.52	4.10	.54	Female
80	2.53	.69	2.68	.47	2.60	.60	Neutral

TABLE 1
Continued

Item Number	Male Group n=28		Female Group n=22		Total Group n=50		Designation
	$\bar{X}m$	SDm	$\bar{X}f$	SDf	$\bar{X}t$	SDt	
81	2.03	.74	2.45	.67	2.22	.73	Ambiguous
82	1.14	.35	1.36	.49	1.24	.43	Male
83	1.35	.67	1.31	.64	1.34	.65	Male
84	3.96	.88	3.77	.68	3.88	.79	Ambiguous
85	1.89	.73	1.90	.52	1.90	.64	Male
86	1.85	.70	2.09	.61	1.96	.66	Male
87	4.42	.74	4.40	.66	4.42	.70	Female
88	3.32	.72	3.09	.29	3.22	.58	Neutral
89	2.21	.83	2.54	.67	2.36	.77	Ambiguous
90	4.39	.91	4.40	.66	4.40	.80	Female
91	2.10	.87	2.59	.59	2.32	.79	Ambiguous
92	2.14	.84	2.54	.50	2.15	.73	Ambiguous
93	4.60	.49	4.81	.39	4.70	.46	Female
94	1.25	.44	1.36	.49	1.30	.46	Male
95	2.92	.26	3.00	.00	2.96	.19	Neutral
96	2.92	.26	2.95	.21	2.94	.23	Neutral
97	4.03	.63	3.86	.35	3.96	.53	Ambiguous
98	4.67	.47	4.68	.47	4.68	.47	Female
99	2.35	.67	2.77	.42	2.54	.61	Neutral
100	1.96	.50	2.31	.47	2.12	.52	Ambiguous
101	3.64	.78	3.63	.49	3.64	.66	Ambiguous
102	2.64	.48	2.27	.45	2.68	.47	Neutral
103	1.53	.57	1.54	.50	1.54	.54	Male
104	1.50	.57	1.22	.42	1.38	.53	Male
105	1.60	.68	1.40	.50	1.52	.61	Male
106	3.57	.74	3.50	.51	3.54	.64	Ambiguous
107	4.46	.57	4.45	.50	4.46	.54	Female
108	3.00	.00	3.00	.00	3.00	.00	Neutral
109	3.00	.00	3.00	.00	3.00	.00	Neutral
110	2.60	.49	2.59	.50	2.60	.49	Neutral
111	1.53	.69	1.86	.35	1.68	.58	Male
112	3.14	.35	3.31	.56	3.22	.46	Neutral
113	4.50	.50	4.63	.49	4.56	.50	Female
114	3.46	.69	3.68	.64	3.56	.67	Ambiguous
115	3.14	.35	3.00	.00	3.08	.27	Neutral
116	1.32	.47	1.27	.45	1.30	.46	Male
117	4.28	.59	4.54	.50	4.40	.57	Female
118	3.17	.77	3.18	.39	3.18	.62	Neutral
119	1.46	.63	1.50	.51	1.48	.57	Male
120	2.96	.18	2.95	.21	2.96	.19	Neutral

varied most from 3.00 would be dropped. Accordingly, the difference between 3.00 and $\bar{X}_t$ was computed for all 36 neutral items, and the 10 items with the highest differences were eliminated. This resulted in a final item pool of 78 items, 26 of which were neutral, 26 male, and 26 female. These items with their sex content judgement ratings are shown in Table 2.

Following the Stage 4 procedure, these 78 items were administered as a test of mathematics problem solving to the remaining 333 subjects. In Stage 5, the results were analyzed in an item analysis routine which computed the point-biserial correlation (r_{pbis}), which is the correlation between a single item and total score, the difficulty level (p), which is the percentage of students answering an item correctly, the Kuder-Richardson 20 reliability coefficient, the test mean, standard deviation, and standard error of measurement. Table 3 contains the r_{pbis} and p values for the 78 items analyzed for the total group of 333 male and female subjects.

The mean performance raw score on this 78 item test for all 333 subjects was 45.82, with a standard deviation of 13.28, and a standard error of measurement of 3.74. The K-R 20 reliability coefficient was .92.

Using the r_{pbis} as the criteria for item selection, the 26 items with the highest r_{pbis} were isolated and labeled Test 1, the Total Group Test. These 26 items with their associated sex judgement rating means and sex content designations are

TABLE 2
78 Items With Sex Content Judgement Ratings
Selected for Final Test

Male Items				Female Items				Neutral Items			
Original Item * #	New Item * #	X	SD_t	Original Item * #	New Item * #	X	SD_t	Original Item * X	New Item * #	X	SD_t
6	73	1.56	.61	8	72	4.46	.61	1	66	2.98	.14
15	24	1.52	.61	12	2	4.46	.57	2	35	3.10	.54
16	38	1.70	.64	23	70	4.40	.76	11	32	2.98	.31
21	57	1.82	.62	25	31	4.50	.73	18	9	2.91	.67
26	26	1.60	.67	27	53	4.36	.74	20	17	2.89	.62
40	15	1.90	.76	32	36	4.16	.76	24	44	3.08	.48
43	63	1.81	.72	34	49	4.18	.82	28	69	2.92	.27
44	11	1.50	.70	35	41	4.26	.85	30	42	2.77	.65
51	1	1.90	.73	38	75	4.34	.68	31	48	3.16	.61
52	23	1.82	.82	41	10	4.30	.70	42	7	2.98	.37
54	62	1.82	.87	47	61	4.30	.61	49	3	2.96	.19
57	16	1.62	.85	55	14	4.53	.68	56	25	2.90	.61
58	20	1.78	.78	59	22	4.14	.78	61	78	3.16	.58
66	34	1.74	.72	60	60	4.24	.79	69	71	2.98	.31
68	54	1.56	.67	62	8	4.18	.62	72	67	3.14	.40
73	65	1.38	.56	64	21	4.26	.69	75	74	3.02	.24
82	18	1.24	.43	65	68	4.26	.85	76	28	3.30	.64
83	52	1.34	.65	74	33	4.32	.85	88	45	3.22	.58
85	55	1.90	.64	79	76	4.10	.54	95	46	2.96	.19
94	13	1.30	.46	87	37	4.42	.70	96	27	2.94	.23
103	59	1.54	.54	90	12	4.40	.80	108	5	3.00	.00
104	40	1.38	.53	93	43	4.70	.46	109	56	3.00	.00
105	50	1.52	.61	98	47	4.68	.47	112	39	3.22	.46
111	58	1.68	.58	107	29	4.46	.54	115	6	3.08	.27
116	19	1.30	.46	117	4	4.40	.57	120	30	2.96	.19
119	77	1.48	.57	113	51	4.56	.50	118	64	3.18	.62

* Original item numbers are those used in the initial pool of 120 items. New item numbers refer to the same items as they were renumbered for inclusion in the final 78 item test.

TABLE 3
Point-biserial Correlation Coefficients (r_{pbis})
and Diffificulty Levels (p) for 78 Item Test, Computed
Using Data From 333 Male and Female Subjects

Item Number	r_{pbis}	p	SJR	Item Number	r_{pbis}	p	SJR
1	.376	.502	1.90	40	.354	.441	1.38
2	.362	.733	4.46	41	.366	.381	4.26
3	.380	.784	2.96	42	.461	.778	2.77
4	.366	.354	4.40	43	.314	.604	4.70
5	.417	.399	3.00	44	.418	.715	3.08
6	.525	.679	3.08	45	.439	.691	3.22
7	.294	.874	2.98	46	.345	.523	2.96
8	.353	.820	4.18	47	.539	.559	4.68
9	.322	.595	2.91	48	.291	.850	3.16
10	.177	.417	4.30	49	.418	.607	4.18
11	.427	.435	1.50	50	.143	.102	1.52
12	.457	.724	4.40	51	.167	.408	4.56
13	.538	.571	1.30	52	.353	.799	1.34
14	.524	.757	4.53	53	.345	.898	4.36
15	.424	.676	1.90	54	.298	.228	1.56
16	.441	.748	1.62	55	.327	.186	1.90
17	.384	.411	2.89	56	.354	.943	3.00
18	.335	715	1.24	57	.259	.294	1.82
19	.187	.462	1.30	58	.435	.706	1.68
20	.398	.619	1.78	59	.502	.571	1.54
21	.238	.943	4.26	60	.306	.901	4.24
22	.396	.658	4.14	61	.335	.754	4.30
23	.373	.694	1.82	62	.429	.664	1.82
24	.213	.402	1.52	63	.396	.312	1.81
25	.182	.177	2.90	64	.164	.709	3.18
26	.469	.730	1.60	65	.455	.438	1.38
27	.471	.676	2.94	66	.428	.826	2.98
28	.469	.742	3.30	67	.477	.336	3.14
29	.396	.658	4.46	68	.418	.607	4.26
30	.408	.727	2.96	69	.400	.706	2.92
31	.380	.550	4.50	70	.209	.285	4.40
32	.480	.607	2.98	71	.411	.751	2.98
33	.371	.559	4.32	72	.434	.405	4.46
34	.335	.553	1.74	73	.427	.652	1.56
35	.454	.760	3.10	74	.490	.456	3.02
36	.226	.345	4.16	75	.321	.357	4.34
37	.409	.712	4.42	76	.343	.465	4.10
38	.491	.607	1.70	77	.262	.402	1.48
39	.472	.580	3.22	78	.557	.565	3.16

displayed in Table 4. The mean performance raw score on this 26 item test for 333 subjects was 16.33, with a standard deviation of 6.07, and a standard error of measurement of 2.09. The K-R 20 reliability coefficient was .88. The mean of the sex judgement ratings for these items was 2.72, putting it in the neutral range, but the sex content designations indicate that 10 of the items were male in content, 4 were female, and 12 were neutral.

The second item analysis was computed using test results for the 185 male subjects alone. The r_{pbis} and p values resulting from analysis of the male data alone are shown in Table 5.

The mean score for the males alone on the 78 items was 47.52, the standard deviation was 12.23, and the standard error of measurement was 3.58. The K-R 20 reliability coefficient was .91.

Again, the r_{pbis} values were used to select the 26 "best" items. The 26 items with the highest point-biserial correlations, their sex judgement ratings and sex content designations are shown in Table 6. These items constitute Test 2, the Male Group Test.

The mean performance raw score on this 26 item Male Group Test for the 333 subjects was 14.56, with a standard deviation of 5.86, and a standard error of measurement of 2.15. The K-R 20 reliability coefficient was .87.

The mean of the sex judgement ratings for these items was 3.08, in the neutral range. The sex content designations were 11 male items, 6 female items, and 9 neutral items.

TABLE 4

26 Selected Items With Highest r_{pbis} from Total Group (N=333) Data, With Corresponding Sex Judgement Ratings (SJR) and Sex Content Designations (SCD)

Item Number	SJR	SCD
6	3.08	Neutral
11	1.50	Male
12	4.40	Female
13	1.30	Male
14	4.53	Female
15	1.90	Male
16	1.62	Male
26	1.60	Male
27	2.94	Neutral
28	3.30	Neutral
32	2.98	Neutral
35	3.02	Neutral
39	3.22	Neutral
42	2.77	Neutral
45	3.22	Neutral
47	4.68	Female
58	1.68	Male
59	1.54	Male
62	1.82	Male
65	1.38	Male
66	2.98	Neutral
67	3.14	Neutral
72	4.46	Female
73	1.56	Male
74	3.02	Neutral
78	3.16	Neutral

TABLE 5
Point-biserial Correlation Coefficients (r_{pbis}) and Difficulty Levels (p) for 78 Item Test Using Data from 185 Males

Item Number	r_{pbis}	p	Item Number	r_{pbis}	p
1	.384	.616	40	.409	.551
2	.400	.670	41	.328	.256
3	.354	.789	42	.406	.822
4	.418	.259	43	.282	.524
5	.380	.395	44	.366	.741
6	.487	.746	45	.399	.735
7	.222	.908	46	.327	.568
8	.307	.789	47	.452	.541
9	.229	.654	48	.247	.892
10	.146	.195	49	.391	.568
11	.513	.665	50	.209	.141
12	.441	.686	51	.084	.222
13	.491	.627	52	.357	.859
14	.479	.757	53	.329	.903
15	.423	.854	54	.376	.281
16	.440	.881	55	.459	.297
17	.349	.373	56	.320	.951
18	.378	.870	57	.295	.395
19	.228	.735	58	.296	.811
20	.428	.757	59	.517	.678
21	.244	.919	60	.304	.897
22	.314	.616	61	.333	.676
23	.305	.838	62	.417	.843
24	.311	.627	63	.509	.389
25	.052	.081	64	.081	.600
26	.401	.822	65	.495	.551
27	.431	.714	66	.444	.816
28	.410	.719	67	.475	.368
29	.355	.562	68	.384	.541
30	.327	.784	69	.372	.741
31	.478	.384	70	.008	.049
32	.436	.649	71	.372	.778
33	.367	.373	72	.440	.243
34	.397	.800	73	.433	.832
35	.452	.800	74	.481	.524
36	.036	.130	75	.298	.314
37	.406	.627	76	.354	.422
38	.424	.778	77	.372	.530
39	.442	.578	78	.551	.643

TABLE 6
26 Selected Items with Highest r_{pbis} from Male Group (n=185) Data, With Corresponding Sex Judgement Ratings (SJR) and Sex Content Designations (SCD)

Item Number	SJR	SCD
4	4.40	Female
6	3.08	Neutral
11	1.50	Male
12	4.40	Female
13	1.30	Male
14	4.53	Female
15	1.90	Male
16	1.62	Male
20	1.78	Male
27	2.94	Neutral
31	4.50	Female
32	2.98	Neutral
35	3.10	Neutral
38	1.70	Male
39	3.22	Neutral
47	4.68	Female
55	1.90	Male
59	1.54	Male
63	1.81	Male
65	1.38	Male
66	2.98	Neutral
67	3.14	Neutral
72	4.46	Female
73	1.56	Male
74	3.02	Neutral
78	3.16	Neutral

The third item analysis was computed using the test results for the 148 female subjects alone. The r_{pbis} and p values resulting from computations using only female data are shown in Table 7.

The mean for the female group on the 78 items was 43.70, the standard deviation was 14.20, and the standard error of measurement was 3.59. The K-R 20 reliability coefficent was .94.

TABLE 7
Point-biserial Correlation Coefficients (r_{pbis}) and Difficulty Levels (p) for 78 Item Test Using Data from 148 Female Subjects

Item Number	r_{pbis}	p	Item Number	r_{pbis}	p
1	.325	.358	40	.248	.304
2	.397	.811	41	.516	.527
3	.413	.777	42	.494	.723
4	.406	.473	43	.434	.703
5	.472	.405	44	.463	.682
6	.539	.595	45	.462	.635
7	.333	.831	46	.347	.466
8	.462	.858	47	.663	.581
9	.391	.520	48	.304	.797
10	.428	.696	49	.493	.655
11	.302	.149	50	.001	.054
12	.528	.770	51	.414	.642
13	.572	.500	52	.321	.723
14	.588	.757	53	.365	.892
15	.394	.453	54	.171	.162
16	.415	.581	55	.020	.047
17	.459	.459	56	.381	.932
18	.258	.520	57	.160	.169
19	.014	.22	58	.520	.574
20	.327	.446	59	.457	.439
21	.322	.973	60	.322	.905
22	.539	.709	61	.457	.851
23	.385	.514	62	.408	.439
24	.055	.122	63	.226	.216
25	.358	.297	64	.417	.845
26	.497	.615	65	.376	.297
27	.500	.628	66	.434	.838
28	.573	.770	67	.474	.297
29	.572	.777	68	.532	.689
30	.458	.655	69	.414	.662
31	.475	.757	70	.568	.581
32	.512	.554	71	.440	.716
33	.636	.791	72	.614	.608
34	.224	.243	73	.389	.426
35	.439	.709	74	.479	.372
36	.588	.615	75	.384	.412
37	.536	.818	76	.373	.520
38	.316	.392	77	.063	.243
39	.517	.581	78	.542	.466

The 26 items with the highest r_{pbis} values, and their sex judgement ratings and sex content designations are shown in Table 8. These 26 items constitute Test 3, the Female Group Test. The mean performance raw score on this 26 item Female Group Test for the 333 subjects was 14.71, with a standard deviation of 5.89, and a standard error of measurement of 2.14. The K-R 20 reliability coefficient was .87. The mean sex judgement rating for these 26 items was 3.64, in the ambiguous range. The sex content designations indicated the selection of 2 male, 14 female, and 10 neutral items.

Table 9 summarizes the out-come of these analyses. It should be noted that the Total Group and Male Group Tests are heavily weighted with non-female items, while the Female Group Test predominates in female items.

As indicated in Stage 6 of the procedure, having constructed three tests, it was of interest to discover the extent of item "overlap" between the tests. To that end, three pairwise comparisons were made. The comparison of the item "overlap" between Tests 1 and 2, the Total Group Test and the Male Group Test, showed that 20 of the 26 items selected as "best" for the Total Group were also selected as "best" for the Male Group. That is, there was a 77% item correspondence between the two tests. Only 23% of the items selected for each test were exclusive to that test. Thus, it can be said that the Male and the Total Group Tests are basically the same test.

TABLE 8

26 Selected Items With Highest r_{pbis} From Female Group Data (n=148), With Corresponding Sex Judgement Ratings (SJR) and Sex Content Designations (SCD)

Item Number	SJR	SCD
5	3.00	Neutral
6	3.08	Neutral
12	4.40	Female
13	1.30	Male
14	4.53	Female
22	4.14	Female
27	2.94	Neutral
28	3.30	Neutral
29	4.46	Female
31	4.50	Female
32	2.98	Neutral
33	4.32	Female
36	4.16	Female
37	4.42	Female
39	3.22	Neutral
41	4.26	Female
42	2.77	Neutral
47	4.68	Female
49	4.18	Female
58	1.68	Male
67	3.14	Neutral
68	4.26	Female
70	4.40	Female
72	4.46	Female
74	3.02	Neutral
78	3.16	Neutral

TABLE 9
Item Sex Content Designations
of Three 26 Item Tests

	Item Sex Content Designation		
	Male	Female	Neutral
Total Group Test	10	4	12
Male Group Test	11	6	9
Female Group Test	2	14	10

The second pairwise comparison involved the overlap between Tests 1 and 3, the Total Group Test and the Female Group Test. Of the 26 items selected for each of these tests, 15 or 58% were the same. Forty-two percent of the items on each test were exclusive to that test. The indication here is that the Total Group Test and the Female Group Test are substantially different.

In the third comparison, items chosen as "best" for the Male Group were compared with those chosen for the Female Group. Thirteen of the 26 items, representing 50% were found on both tests, leaving the remaining 50% exclusive to each test. Like the Female/Total Group Test Comparison, these two tests are basically different.

The overlapping items, and the percentage of each test comparison that they represent are shown in Table 10.

Based on the definition of "bias" presented earlier, it can be said that the comparison of the Female Group Test with the Male Group Test showed the highest percentage of biased items. In that comparison, 50% of the items selected for each test were exclusive to that test. The interaction of group membership and items selected is highest in that instance.

The comparison of the Total Group and Male Group Tests showed the least difference in the items selected. 77% of the items selected for the Total Group Test were the same as those selected for the Male Group Test.

TABLE 10
"Overlapping" Item Comparisons on Three Tests

"Overlapping" Items		
Total Group Test vs. Male Group Test	Total Group Test vs. Female Group Test	Male Group Test vs. Female Group Test
6	6	6
11	12	12
12	13	13
13	14	14
14	27	27
15	28	31
16	32	32
27	39	39
32	42	47
35	47	67
39	58	72
47	67	74
59	72	78
65	74	
66	78	
67		
72		
73		
74		
78		
Percentage Represented by "Overlapping" Items		
.77	.58	.50

In Stage 7 of the data analysis, the mean performance scores and associated standard deviations for males and females on each of the three constructed tests were computed. Because all of the 333 subjects had completed all 78 items from which the test had been constructed, it was possible to compute the average performance for females on the items selected as "best" for the female group, and also on items selected as "best" for the male group, and for the total group. These same means were computed for the male subjects. These means and standard deviations are shown in Table 11.

Two analyses of variance were computed, the first comparing the means for the female group on the three tests for significance of differences, and the second comparing the male means in the same way. The analysis of variance of female means is shown in Table 12, and the same information for the male comparison is shown in Table 13. The F value in the comparison of female means on the four tests was 169.45 while the F statistic for the comparison of male means was 176.64. Both of these values are significant at $\alpha = .01$.

In order to identify which of the means were different specifically, the Tukey a posteriori test was computed separately on the means for males and females on the three tests. The Tukey test on female means resulted in significant differences at the .99 level of significance for all pairwise comparisons. The same results (i.e., all pairwise comparisons produced significant differences) was found for male means. Computation of these a posteriori test can be found in Appendix V.

TABLE 11
Mean Performance Scores for
Males and Females on Three Tests*

	Male Mean n_m = 185	S.D.	Female Mean n_f = 148	S.D.
Test 1 Total Group Test	17.90	5.38	14.36	6.34
Test 2 Male Group Test	16.02	5.66	12.93	5.64
Test 3 Female Group Test	13.98	5.20	15.61	6.49

* The number of items omitted by the two sex subgroups was computed and was found not to be significantly different.

TABLE 12
Analysis of Variance of Female Means
on Three Constructed Tests

Source of Variation	SS	df	MS	F
Between People	16,288.68	147		
Within People	981.67	296		
Treatment	525.33	2	262.66	169.45*
Residual	456.65	294	1.55	
Total	17,210.35	443		

*$F_{.99}$ $(2, \infty) = 4.61$

TABLE 13
Analysis of Variance of Male Means
on Three Constructed Tests

Source of Variation	SS	df	MS	F
Between People	16,168.87	184		
Within People	3,061.67	370		
Treatment	1,497.95	2	748.97	176.64*
Residual	1,563.72	368	4.25	
Total	19,230.54	554		

$^*F_{.99}\ (2,\infty) = 4.61$

A number of Analyses of Covariance (ANACOVA) were computed between male and female means on the three tests. Research cited earlier (Berry, 1958; Aiken and Dreger, 1961; Aiken, 1963; Brown and Abel, 1965; etc.) has confirmed the existence of a relationship between attitude toward mathematics and achievement in that subject. Findings indicate that males have more favorable attitudes towards mathematics, and that this is a contributing factor in their superior achievement. In order to compare male and female achievement on the three tests constructed in this study, without the confounding influence of attitude, ANACOVA were computed with subjects' scores on the Mathematics Attitude Scale (Aiken and Dreger, 1961) as the covariate.

In the first ANACOVA, male and female scores on the Total Group Test, where the male group scored higher, were compared. The resulting F statistic was 15.941 with 1 and 330 degrees of freedom, significant at the .01 level of significance. In the second ANACOVA, scores on the Male Group Test, where males had also scored higher, were compared, and an F of 11.780 resulted, also significant at $\alpha = .01$. The third ANACOVA, a comparison of scores on the Female Group Test, the only test on which females scored higher, produced an F statistic of 18.038, which is also significant at $\alpha = .01$. It is important to note that the first two ANACOVA's indicated that on the Total Group and Male Group Tests males scored significantly higher than females even after removing the effect of attitude toward mathematics (as measured by the Mathematics Attitude Scale). On the third AVACOVA, however, where females had scored significantly higher, the significance of the difference in favor of the females still remained after controlling for differences in attitude.

The second set of ANACOVA's were computed using the subjects' Verbal score on the Differential Aptitude Test as the covariate. Research by Spilman and Weiner (1972) has shown that differences in verbal ability between males and females is a possible confounding factor in mathematics achievement. Accordingly, the influence of verbal ability was covaried in analysis of male and female achievement on the three tests. The ANACOVA for the Total Group Test resulted in an F of 43.862, for the Male Group Test the F was 33.712, and for the Female Group Test an F of 9.938 resulted. All three F statistics are significant, the first two in favor of the males, but the third in favor of the females, indicating that removal of the possible confounding influence of verbal ability does not remove the significant differences in performance between the sexes.

The final test computed was an ANACOVA on the Total Group Test scores for males and females using two covariates, both Mathematics Attitude Scale score, and Verbal score on the Differential Aptitude Test. This ANACOVA resulted in an F of 11.024 which is also significant at $\alpha = .01$ indicating that controlling both Mathematics Attitude and Verbal Ability still does not reduce the difference in performance between male and female subjects.

V DISCUSSION AND CONCLUSIONS

Two possible sources of bias in tests of mathematical problem solving were explored in this research. The first had to do with the bias caused by the unequal familiarity of the sexes with the content of the items. The second was the bias resulting from the commonly employed practice of selecting items from an item pool using a mixed sex group of subjects. Three hypotheses were presented:

1) If "best" items are selected from an item pool separately for sex subgroups, these items will vary from those selected as "best" for the total group.

2) The variation in items will be partially explained by the differential familiarity of item content to the subgroups.

3) Subgroup performance on "best" items selected using that subgroup as the tryout sample will be significantly higher than their performance on items selected for the other subgroup, or for the total group.

In order to explore the outcomes of this research, let us consider the implications of the results of the pertinent stages.

Stage 3 of the study involved student ratings of the original 120 item pool as to the differential familiarity of the item content to male and female high school students.

The directions to the students required that they read the items not with a view toward solving the problems, but rather, concentrate their attention on the content in which the problems were imbedded. They were asked to decide, using a 1-5 Likert Scale, whether the item's content was likely to be more familiar to males or females. It is of interest here to point out that while none of these students had ever been asked to view mathematics problems in this way, there was ready acceptance of the notion, and the students had no difficulty in designating items as male, female, or neutral. The results followed what might be viewed as a perpetuation of our cultural stereotyping. Those items designated as female had a tendency to be concerned with "typically female" pursuits: cooking, sewing, planning parties, picnics, luncheons. Male items followed the same pattern and included areas such as: sports, automobiles, business careers, building, etc. Neutral items commonly had little, if any, content. That is, an item such as:

$$r = 18$$
$$r/s = 6/y$$
$$y/s = ?$$

was viewed as neutral. These findings were an extension of findings in the Pilot Study which had preceded this research, and are indicative of the fact that the tendency to view pursuits stereotypically where sex role is involved is a continuing problem with the present generation of high school students.

Stage 5 of the study was concerned with an exploration of the first two hypotheses. Three test construction procedures were employed which varied only in the make-up of the population used in the item analyses. First, a group of 26 "best" items, where "best" was defined as items having the highest point-biserial correlations, were selected from an item analysis conducted on the data of all 333 male and female subjects. This test, labeled the Total Group Test, was found to have a mean difficulty level of .628, for the total group, a mean sex-judgement rating of the items of 2.72, and an item sex designation make-up of 12 neutral, 10 male, and 4 female items. Remembering that the range of neutrality for items was 2.5-3.5, it appears that this test is neutral in sex content. The count of the sex designations of the items, however, indicated that there were a very small number of female items included. Of the 26 selected items, only 15% were female.

The second item analysis was computed to select the 26 "best" items when the subject population was made up only of male students. In this second case the test had a mean difficulty level of .563 for the total group, a mean sex judgement rating of 3.08, and an item sex designation of 9 neutral, 11 male, and 6 female items. Here, too, the test appears to be neutral in content, but a review of the item types shows that on this Male Group Test only 27% of the selected items were female, while 42% were male.

The third analysis selected the 26 "best" items using the data for the female subjects alone, and produced a test with a mean difficulty level of .573, for the total group, a mean sex judgement rating of 3.64 and an item sex content make-up of 10 neutral, 2 male, and 14 female items. This third test, therefore, contained 54% female items.

A review of the extent of actual item overlap (i.e., How many items appear on the Total Group Test and also on the Male Group Test, etc.), as indicated in Stage 6, produced the finding that there is a 77% correspondence between items appearing on the Total Group Test and those on the Male Group Test, a 58% correspondence between items on the Total Group Test and Female Group Test, and only a 50% correspondence between items on the Male and Female Tests.

These results indicate a confirmation of the first two hypotheses. When "best" items are selected separately for males and females, 50% of the items vary on the two tests. Further, the Female Test also varies substantially from the test selected for the total group, having only a 58% item correspondence. It is interesting to note, however, that the Male Test does not show as great a variation with the Total Group Test, having a 77% item correspondence. It appears that the Male Test and the Total Group Test are substantially the same. It is also true that the variation in items selected for the three tests is at least partially explained by the differential familiarity of the item content to the subgroups. While the mean sex judgement ratings for all three tests fell within the neutral range. The sex designations of the items selected for each test were quite different. The Total Group and the Male Tests contained few female items, but the majority of the items on the Female Test were female in content. The implication here is that female items are "better" for females than male or neutral items. Fifty-four percent of the items selected for the Female Test were female, but the remaining selected items were not evenly divided between neutral and male. It would appear that neutral items are substantially "better" than male items for a female population, as 38% of the Female Test was neutral in content, while only 8% was male.

In Stage 7, the mean performance scores for males and females on each of the three tests were computed. The results showed females scoring highest on the Female Test. That is, the test composed of the 26 items found to be "best" for females resulted in a mean score which was significantly higher than their mean score on the Male Group Test or on the Total Group Test. Further, the between group comparisons of male and female performance showed females scoring significantly higher than males on this test constructed specifically for their subgroup. Reviewing male performance on the three tests it was found that they scored highest on the Total Group Test (which contained the fewest female items), and significantly least well on the Female Group Test. The between group analysis showed males scoring significantly below females only on the Female Group Test.

The picture emerges of females scoring best on their own test, second best on the test constructed using the total group. Males do best on the Total Group Test, second best on their own test, and least well on the Female Group Test. That is, it is found that females achieve the highest mean score when the test is heavily weighted (54%) with female items, and least well when the test has few female items (27%), but many male items.

In terms of Green's (1972) findings, presented earlier, this outcome is significant. Green had found that racial sub-groups scored higher on tests made up of items selected using that subgroup alone as the tryout sample. The same outcome eventuates here, where sex subgroups are used separately as the tryout samples. When female item data alone is used to select "best" items for a mathematics problem solving test, the female subgroup scores higher on that test. It might be suggested that this test is, in fact, easier, and that would account for the higher scores achieved by females. This possibility, however, was considered and precautions taken against it. It should be recalled that the original items were written with a concern for balance between male, female, and neutral items in terms of item types (i.e., number of problems involving basic arithmetic functions, percents, decimals, fractions, algbraic functions, etc.), and difficulty of items. The similarity of items in these respects was verified by a panel of high school mathematics teachers. Further, the difficulty level (i.e., p value) of the items was computed in the original item analysis and that p value for the total group (i.e., N= 333) recorded. In order to verify the fact that the Female Group Test was not easier, the p values for items selected for that test (using the total group item analysis results) were averaged. The same averages were computed for the Total Group Test items, and the Male Group Test items. The average difficulty level of the items on the Female Group Test was .573, while the average for the Total Group Test was .628, and for the Male Group Test, .563. The p value, or difficulty level statistic, represents the percentage of students who got that item correct. Therefore, the figures presented here represent the approximate average difficulty of the overall test. The .628 figure indicates, for example, that, on the average, 63% of the students (N=333) answered the items on the Total Group Test correctly, thereby making that the easiest test. Fifty-seven percent of the students got the Female Group Test items correct, while 56% got the Male Group Test items correct. It would appear that there was no significant difference in difficulty of items between the three tests. In addition, while the Female Group Test resulted in the highest mean performance scores for females, this was not true for males. Males scored highest on the Total Group Test, and <u>least well</u> on the Female Group Test. Obviously, the difficulty level of the items is <u>not</u> the overriding factor in determination of sex subgroup performance. The factor which does seem, however, to be contributing to performance scores is the number of items, within a test, of greater content familiarity to one sex group or another.

The Total Group Test, on which males scored highest, contains the fewest female content items (i.e., four). The Female Group Test, on which males scored least well, contains the highest number of female items, 14. Conversely, females scored least well on the Male Group Test which contains the most male items, 11, and best on the Female Group Test which contains the fewest male items, 2. It is apparent that the differential familiarity of the content of the items to the two sex groups plays a part in determination of their performance.

In comparing male and female achievement between groups, as opposed to within groups, an analysis of covariance technique was used. This procedure was used based on the confirmation of the existence of strong relationships between mathematics performance and several other factors. The relationship between performance and attitude has been well documented, and accordingly, attitude toward mathematics for all subjects was measured and covaried in the between group analyses. This control over attitude, however, did not lessen the significance of the differences between male and female subgroups in performance. Males continued to achieve significantly higher than females on the Male Group and the Total Group Tests, but significantly lower than females on the Female Group Test. This last outcome is particularly important in light of the traditional superiority of males on tests of mathematics problem solving.

On this test, developed using a female tryout sample, males do significantly less well than females. When one considers the fact that the test has been shown to be no easier than the other tests constructed, and that the item types do not vary from other problem solving tests, it is of considerable note that females score significantly higher on it. In addition, it remains clear, from this result, that even the removal of the students' attitudes toward the subject matter of the test, does not lessen the effect of the differential familiarity of the item content to the sexes.

The same result was found when verbal ability was controlled using an analysis of covariance to compare male and female performance. Differences between the sex subgroups in verbal ability are not responsible for differences in performance. Controlling this factor males continued to score significantly higher than females on the Male Group and Total Group Tests, while females scored significantly higher on the Female Group Test.

The third hypothesis put forth in this research has, in part, been verified. Female subgroup performance on "best" items selected using that subgroup alone as the tryout sample, is significantly higher than their performance on items selected for the other subgroup, or for the total group. The exception to the hypothesis is for male subjects who score highest on the test constructed using the total group as the item tryout sample. This, however, contributes more evidence to the finding that it is the differential familiarity of the content of the items to the sexes which is the most important factor in performance. It must be remembered that the Total Group Test was more "non-female" in content than was the Male Group Test, the former having only four, or 15% female items, while the latter had 6, or 23% female items. The only test on which they score significantly less well is the Female Group Test which contains 54% female items.

The objective of this research was to document "bias" toward females in the content of test items, and to show that females would score higher on items which are not "biased", although similar in subject matter and difficulty, and it appears to have been proven. The difficulty remaining stems from the temptation to recommend to test makers that they develop tests separately for female sex subgroups. This approach to test construction and administration is unrealistically costly, and further, introduces its own kind of bias. It should not be necessary to treat females "differently" in order to maximize their performance on mathematics problem solving tests. One of the original thoughts behind this research has been the hope that presentation of a balanced item pool, in terms of the differential familiarity of the items to the sexes, might result in a test on which female performance would be maximized, and the difference between males and females in performance would be lessened. If it is remembered that, traditionally, mathematics problem solving items are overwhelmingly male in content familiarity starting even with the test construction item pool, then it can be seen that it was thought possible that by balancing the item pool (i.e., even numbers of male, female, and neutral items) the potential existed for selection of "best" items which also represented a balance. Interestingly, this was not the case. The "best" items selected for the total group were, instead, ten male, twelve neutral, and only four female. This outcome apparently stems from the very notion underlying the point-biserial correlation (r_{pbis}).

The r_{pbis} selects items which discriminate effectively between high and low scoring students. That is, if a student achieves a high overall score, one should expect, with high probability, that he/she would answer any given item correctly. On the other hand, if a student achieves a low overall score, one expects that he/she would answer any given item incorrectly. Based on this idea, a high point-biserial correlation results for items which high scoring students get right, and low scoring students get wrong, thereby serving as an efficient discriminator between the two groups. In the case of mathematics problem solving tests, if one were to list the students by sex in descending order of score, one would discover that the high scorers are predominantly male, while the low scorers are predominantly female.* The r_{pbis}, therefore, is effectively selecting as

* In the research reported here the highest scoring quarter of the students was made up of 64 male and 19 female students, while the lowest quarter had 57 females and 26 males.

"best" those items which males answer correctly, and females answer incorrectly. Because, even with a balanced item pool, females continue to achieve lower scores on the overall test, this artifact of the statistic is not overcome, and the selected items, although not more difficult, are those which females are more likely to complete unsuccessfully. If we can assume that the distribution of mathematics ability is the same for females as it is for males then this statistic is not discriminating effectively between them. In effect, all females are "lumped" into the low scorer category and items which high scoring females answer correctly are not chosen as "best" items for a test. This artifact is maximized when test items are all male in content. It is reduced when the item pool is balanced in content, but the reduction is not sufficient to allow females to achieve their maximum scores.

Rather than opting for separate test construction and administration procedures for sex subgroups, it was considered that item selection techniques might be varied. At present, "best" items are selected, as they were here for the Total Group Test, without regard for the differential familiarity of the item to the sexes. The number of items with the highest r_{pbis}'s needed are simply extracted and combined into a test. If, instead, "best" items were selected separately from among those male in content, then female in content, and then neutral, the test would necessarily be balanced in regard to familiarity. To discover the effect of this procedure on performance this was done with the items used in this test construction. Using the statistics computed on the total group (N=333) on the total item pool (i.e., 78 items), the nine "best" male items, nine "best" female items, and the eight "best" neutral items were selected and assembled. The difficulty levels of each item for males alone (n_m=185), and females alone (n_f=148) were recorded and summed. (The sum of the item difficulty levels for a group equals the mean score for that group on that selection of items.) This information is shown in Table 14. The results indicate that on this stratified sample of items both males and females would score higher, on the average, then on any of the previously constructed tests. The question of interest, however, is whether the difference between male and female performance is reduced by use of this technique. Calculation of differences between the means for males and females on each of the tests is shown in Table 15. It can be seen that the use of a stratified sample of items does reduce the difference between male and female performance, although not dramatically. It may be, that from a test construction point of view, this is the fairest procedure that can be used at present.

One other possibility suggested by Donlan (71) was explored here. If only neutral mathematics items were included in a test, Donlan had found that the difference in performance between males and females was considerably

TABLE 14

Items Selected Using Stratified Sampling Technique, with Associated Means for Males and Females

Item	p_m n_m=185	p_f n_f=148
11	.665	.149
12	.686	.770
13	.627	.500
14	.757	.757
15	.854	.453
16	.881	.581
22	.616	.709
26	.822	.615
27	.714	.628
28	.719	.770
30	.784	.665
35	.800	.709
37	.627	.818
42	.822	.723
44	.741	.682
45	.735	.635
47	.541	.581
49	.568	.655
58	.811	.574
59	.678	.439
62	.843	.439
65	.551	.297
66	.816	.838
68	.541	.689
72	.243	.608
73	.832	.426
	$\bar{X}_m$ 18.27	$\bar{X}_f$ 15.71

TABLE 15

Mean Performance Scores and Differences for Males and Females on Four Tests

	Male Mean n_m=185	Female Mean n_f=148	Difference
Total Group Test	17.90	14.36	-3.54
Male Group Test	16.02	12.93	-3.09
Female Group Test	13.98	15.61	+1.63
Stratified Sample Test	18.27	15.71	-2.56

reduced. Accordingly, the 26 neutral items were separated from the 78 item pool, and the difficulty level on each item for each sex subgroup was identified. The difficulty levels were then summed to find the mean performance score for that sex subgroup on this neutral item test. The results of this analysis are shown in Tables 16 and 17. On this neutral test the male mean is 17.37 while the female mean is 16.21. The difference between the means is 1.16. Comparing this difference to those displayed in Table 15, one can see that it is the smallest of the differences computed. It must be remembered that these neutral test means were computed on the same items for the two sex subgroups. This finding might seem to support the idea that mathematics tests should be neutral in content. This is not the case, however, as neutral items are not completely representative of mathematics problem solving items. They do not require the application of mathematical strategies to real life problems, as they tend to emphasize only the computational aspect of the problem. The fact that males and females score more similarly on these items only provides further evidence that the content of items is a crucial factor in determination of how a student will perform.

In summary, this research strongly suggests that the differential familiarity of mathematics problem solving items to sex subgroups <u>is</u> a factor which effects performance. This factor is a consideration for both males and females.

TABLE 16

Difficulty Level for Males (n_m= 185) on 26 Neutral Items

Items	P Males
3	.789
5	.395
6	.746
7	.908
9	.654
17	.373
25	.081
27	.714
28	.719
30	.784
32	.649
35	.800
39	.578
42	.822
44	.741
45	.735
46	.568
48	.892
56	.951
64	.600
66	.816
67	.368
69	.741
71	.778
74	.524
78	.643

$\overline{X}_m$ = 17.37
Neutral
Items

TABLE 17

Difficulty Level for Females (n_f= 148)
on 26 Neutral Items

Item	P Males
3	.777
5	.405
6	.595
7	.831
9	.520
17	.459
25	.297
27	.628
28	.770
30	.655
32	.554
35	.709
39	.581
42	.723
44	.682
45	.635
46	.466
48	.797
56	.932
64	.845
66	.838
67	.297
69	.662
71	.716
74	.372
78	.466

$\overline{X}_f$ = 16.21
Neutral
Items

Females do not deal effectively with problems which are heavily male in content, nor do males deal well with problems heavily female in content, regardless of the overall difficulty of the items. This fact may have been overlooked in previous research because few, if any, tests have been constructed which utilized an equal number of male, female, and neutral items, and items for mathematics tests have not earlier been selected separately for sex subgroups.

Further, the outcomes of this research indicate that the commonly used point-biserial correlation introduces its own bias. In a situation where one subgroup scores predominantly in the upper levels of those tested, with the other predominating in the lower levels, the items selected as "best" may be unfair to the lower scoring group. This outcome also eventuates regardless of the difficulty levels of the items in question.

It would seem that the best technique, at present, for construction of unbiased tests of mathematics problem solving, is the initial use of a balanced item pool (i.e., equal numbers of male, female, and neutral items), and selection of "best" items based on a point-biserial correlation, but using a stratified sampling technique. That is, a balanced number of "best" items would be selected from among those categorized as male, female, or neutral in terms of content familiarity to the sexes.

In order to better understand the outcomes of this research a number of areas seem worthy of further investigation.

The present study was carried out with subjects in Grades 10-12, where sex differences in performance on mathematics problem solving tests are both pronounced and consistent. It would be of interest to conduct similar research using a developmental approach. Tittle, McCarthy, and Steckler (1973) have documented the fact that mathematics tests used in the lower grades also contain item content sex bias. The effect of elimination of this bias on students in lower grades, where sex differences in performance are less outstanding, would be of interest.

Further, the predictive validity studies of the existing measures for high school students sadly lack specificity. For example, validity studies of the S.A.T. mathematics for females alone do not appear to exist. Studies are needed which establish the predictive validity of the test for female students enrolled in required college mathematics courses. It would then be of interest to discover if tests developed on sex subgroups alone, or those using stratified sampling techniques produced an increase in predictive validity.

There is no question but that the present research further substantiates the existence of sex differences, at least in the area of performance on mathematics problem solving measures. It also indicates that one cause of the differences is the inability of sex subgroups to deal with item content with which they are unfamiliar. Minuchin (1972) has outlined goals for education which include the minimization of sex stereotyping; the exposure of females to a broader range of experiences, ideas, and female models; a concerted effort to educate females in skills of choice, problem solving and evaluation; and an encouragement of female development in self-differentiation and self-knowledge. These goals are slowly being adopted by schools, and their effect, with the passage of time, should minimize sex differences in performance. The effects of these changes need to be carefully, and consistently documented. Hopefully, the problem of differential familiarity of subject matter content will be reduced with the introduction of educational change, and the elimination of stereotyping, and this reduction will need to be closely monitored.

APPENDIX I

PLEASE PROVIDE THE INFORMATION ASKED FOR BELOW:

1. Birth date ______________________________

2. Name ______________________________

3. Grade in high school (circle one) 10th 11th 12th

4. Sex (circle one) Male Female

5. Do you plan to attend college (circle one) Yes No

6. How would you rate your ability in mathematics

1	2	3	4	5
Poor		Average		Excellent

DIRECTIONS:

Please read each of the items on the following pages. As you read them, do not focus your attention of the mathematics involved, but rather, on the content of the item. After reading an item decide whether you believe that the content would, in general, be more familiar to males or females. Judge each item on the sliding scale provided, where 1 would indicate that that item has content which would be much more familiar to males than females, and 5 would indicate content much more familiar to females than males. The males and females for whom you are judging are your age, but remember that you're judging for the <u>average</u> male or female high school student.

1	2	3	4	5
Very Familiar to Males		Equally Familiar to Males & Females		Very Familiar to Females

1. $\frac{1/3 - 1/12}{1/4} =$

   1 2 3 4 5
   male female

2. There are 45 pupils in a certain class. If 2/3 of the pupils are boys, and 1/2 of the boys have blue eyes, how many blue eyed boys are in the class?

   1 2 3 4 5
   male female

3. If 3 secretaries can type 6 manuscripts in 12 days, how many days would it take 2 secretaries to type 3 such manuscripts?

   1 2 3 4 5
   male female

4. A can of food feeds 3 kittens or 2 adult cats. If I had 8 cans of food, and I feed 12 kittens, how many adult cats can I feed with the remainder?

   1 2 3 4 5
   male female

5. Mr. A owes Mr. B $70.00, and Mr. B. owes Mr. A $60.00. If Mr. A gives Mr. B a fifty dollar bill, how many dollars in change should Mr. B give Mr. A?

   1 2 3 4 5
   male female

6. One gear rotates once every 7 minutes, and another rotates once every 5 minutes. Both gears begin to rotate at the same time every (?) minutes.

   1 2 3 4 5
   male female

7. 234, 256, 273, 281, 218X
Of the numbers listed above (assuming that X is greater than 1), which of the following numbers cannot possibly be the average of the five numbers?

1 2 3 4 5
male female

8. In making drapes, Joan needs to cut 8 foot panels. How many panels may she cut from fabric that is y yards long?

1 2 3 4 5
male female

9. Mr. Berg, who works twice as fast as Mr. Slocum, receives an hourly rate of pay 1-1/2 times as much as Mr. Slocum. An efficiency expert calculates that an article produced by Mr. Berg has a labor cost of 12¢. What is the labor cost of an article produced by Mr. Slocum?

1 2 3 4 5
male female

10. Mary purchases 3 pounds of mixed nuts at 89¢ per pound. Jan purchases 2 pounds of peanuts at 49¢ per pound. What is the average price per pound for the mixture of both purchases?

1 2 3 4 5
male female

11. A student reads at the rate of 50 pages per hour. How many minutes will it take him to read 2/5 of a page at this rate?

1 2 3 4 5
male female

12. A recipe for cheesecake calls for 12 ounces of cottage cheese and 6 ounces of cream cheese. If there are 216 calories in 8 ounces of cottage cheese, and 106 calories in 1 ounce of cream cheese, how many calories are there in 36 ounces of the cheese mixture?

1 2 3 4 5
male female

13. A dietician has sufficient milk to feed 13 infants for 4 weeks. How many days will this supply last if 13 additional infants are added?

1	2	3	4	5
male			female	

14. A typist can complete a task in 3 hours. What part of her job can she do from 8:55 to 9:15?

1	2	3	4	5
male			female	

15. A mile runner is clocked at 58 seconds after the first quarter mile (440 yards), and 1 minute 56 seconds after he has run a half mile. What rate (in yards per second) must he maintain for the final two quarters if he is to run a 4 minute mile?

1	2	3	4	5
male			female	

16. A carpenter is supposed to build a minimum of t tables. In d days he builds m more tables than the minimum. What is the average number of tables he produced each day?

1	2	3	4	5
male			female	

17. A merchant paid $30.00 for an article. He wishes to place a price tag on it so that he can offer a 10% discount on the price marked on the tag, and still make a profit of 20% on the cost. What price should he mark on the tag?

1	2	3	4	5
male			female	

18. A corporation has 8 departments with 10-16 bureaus each. In each bureau there are at least 40, but no more than 60 workers. If 10% of the workers in each bureau are typists, what is the minimum number of typists in a department?

1	2	3	4	5
male			female	

19. If 2/3 of the men in a factory go on vacation in July, and 1/2 of the remainder take their vacations in August, what fraction of the men take their vacations at other times of the year?

1	2	3	4	5
male			female	

20. A picture in an art museum is six feet wide and eight feet long. If its frame has a width of 6 inches, what is the ratio of the area of the frame to the area of the picture?

1	2	3	4	5
male			female	

21. If 8 men can build a house in 12 days, what is the percentage increase in the number of days required to do the job if 2 men are laid off?

1	2	3	4	5
male			female	

22. One-tenth of a man's salary is spent for clothing, one-third for food, and one-fifth for rent. What percentage of the salary is left for other expenditures and savings?

1	2	3	4	5
male			female	

23. The batter which is 7 inches high in a loaf pan 1-1/4 feet by 8 inches, is poured into a pan 13 inches by 20 inches. What height will it reach in the larger pan?

1	2	3	4	5
male			female	

24. In a library system having six branches there are 60 workers employed. If no library has less than 7 workers and no more than 18, what is the minimum number of workers in any two of these branches?

1	2	3	4	5
male			female	

25. Jane cut a length of cloth 13 yards, 5 feet, and 1 inch into three equal parts. The length of each piece is?

1 2 3 4 5
male female

26. An architect uses a scale of 1/4 inch to a foot. If the dimensions of a roome as represented on the plans of a house are 4-1/2 inches by 3-3/4 inches, the actual dimensions of the room in feet are?

1 2 3 4 5
male female

27. In making tuna fish salad, a recipe calls for 2 cups of fish to 1/8 of a cup of chopped celery. How much celery should be used for 8 cups of tuna fish?

1 2 3 4 5
male female

28. If X/16 = .374, what is the numerical value of X?

1 2 3 4 5
male female

29. There are 120 children in a nursery school that receive a pint of milk daily. How many 5 gallon cans of milk are used each day?

1 2 3 4 5
male female

30. If 2 parts of sand are mixed with 3 parts of gravel, what part of the total mixture is sand?

1 2 3 4 5
male female

31. Ten broken hens' eggs wigh one pound. What is the weight, in ounces, of 1 broken hen's egg?

1 2 3 4 5
male female

32. A secretary types business letters at a rate of d words per minute, and statistical tables at a rate of u words per minute. What is her average rate of typing if she types for 2 hours?

1	2	3	4	5
males			female	

33. Sixteen ounces of fresh orange juice contains 216 calories, and 16 ounces of fresh grapefruit juice contains 174 calories. If an 8 ounce mixture of these two juices contains 94 calories, what fraction of the mixture is orange juice?

1	2	3	4	5
male			female	

34. A department store had an end-of-winter sale on ladies coats at the following prices:
 A- \$80. coats reduced to \$55.
 B- \$85. coats reduced to \$60.
 C- \$90. coats reduced to \$65.
 D- \$95. coats reduced to \$70.
 E- \$120. coats reduced to \$95.
Which group of coats was offered at the greatest rate of discount from its original price?

1	2	3	4	5
male			female	

35. In planning the annual Girl Scout Jamboree, Mrs. Smith donates 15 pounds of hot dogs, and Mrs. Jones donates 9 pounds. Together they represent 16-2/3% of the total number of pounds needed. How many more pounds must be donated to assure the organizer that at least .75 of the amount needed is available?

1	2	3	4	5
male			female	

36. A chemist has 80 pints of a 20% salt solution. How much pure salt must he add to produce a solution which is 30% pure salt?

1	2	3	4	5
male			female	

37. The state of Oklahoma had 8 Congressmen in 1940, and 6 in 1950. What was the percent change in their representation?

1	2	3	4	5
male			female	

38. If 2s seamstresses can sew 2d dresses in 2w weeks, how many seamstresses would it take to sew 4d dresses in 4w weeks?

1	2	3	4	5
male			female	

39. A man binds 112 books in one day, and his assistant works 1/4 as fast. How many days would it take them to bind 560 books if each works alone on alternate days?

1	2	3	4	5
male			female	

40. A motorist leaves at 9:00 AM and stops at 9:20 AM. If he covered 18 miles, what was his average velocity for this part of the trip?

1	2	3	4	5
male			female	

41. Mrs. Crocker can do the dinner dishes in a quarter of an hour, while her daughter takes half an hour. What part of an hour would it take to do the dishes if they work together?

1	2	3	4	5
male			female	

42. If 2 halves of 2-1/2 is added to 2-1/2 the result is?

1	2	3	4	5
male			female	

43. A man runs y yards in m minutes. What is his rate in yards per hour?

1	2	3	4	5
male			female	

44. Ten merchants agreed to purchase uniforms for a local baseball team at a total cost of m dollars. After a disagreement, 2 merchants dropped out of the project. By how many dollars was the cost to each remaining merchant increased?

1 2 3 4 5
male female

45. On the average an inspector rejects .08% of the medical instruments produced in a factory as defective. How many medical instruments will he inspect in order to reject 2?

1 2 3 4 5
male female

46. A ranch is surrounded by a fence 320 feet long having wooden posts each 40 feet apart. How many posts are there?

1 2 3 4 5
male female

47. One cup of condensed, sweetened milk weighs 11 ounces. How much milk will remain unused for a recipe requiring 2 cups of milk when a housewife opens 4 six ounce cans of condensed milk?

1 2 3 4 5
male female

48. A trailer carries 3,4 or 5 crates on a trip. Each crate weighs no less than 125 pounds and no more than 250 pounds. What is the minimum weight (in pounds) of the crates on a single trip?

1 2 3 4 5
male female

49. r = 18
r/y = 6/y
y/s = ?

1 2 3 4 5
male female

50. A nurse gives her patient one pill every 45 minutes. How many pills will she need for her 9 hour tour of duty if she gives the patient a pill at the beginning and end of her tour?

1 2 3 4 5
male female

51. A fishing rod manufacturer finds that 0.4% of his production is defective and not suitable for marketing. How many rods of each 1000 produced will be rejected?

1 2 3 4 5
male female

52. A car uses a gallon of gasoline in traveling 15 miles. Another automobile can travel m miles on a gallon of gasoline. How many miles can the second travel on the amount of gasoline required by the first car in going 60 miles?

1 2 3 4 5
male female

53. How long is the shadow of a 35 foot oil rig, if a 98 foot oil rig casts a 42 foot shadow at the same time?

1 2 3 4 5
male female

54. Seven type p nails are worth 5 type q nails. Three type p nails are worth x type q nails. What is the numerical value of x?

1 2 3 4 5
male female

55. Sue bought 8 feet of ribbon to use to decorate baskets. How many 1/4 inch strips will she be able to cut?

1 2 3 4 5
male female

56. The Japanese ken is equivalent to 5.97 feet. How many feet are there in 59.7 ken?

1	2	3	4	5
male			female	

57. After several tryouts, 20% of a football squad was discharged. The coach then had 32 players. How many players were on the squad at first?

1	2	3	4	5
male			female	

58. During the month of June, the General Motors Corporation sold twice as many Chevelles as Impalas, 3 times as many Vegas as Chevelles, and 4 tiems as many Novas as Vegas. If they sold 300 more Vegas than Chevelles, how many Novas were sold?

1	2	3	4	5
male			female	

59. In planning for the church picnic, the ladies guild purchases enough hot dogs to allow each of the 18 person 3 hot dogs. If 9 additional persons come to the picnic, how many fewer hot dogs will each person be given?

1	2	3	4	5
male			female	

60. A woman purchases 4 pounds of chopped meat at $.80 per pound. What change does she receive from a ten dollar bill?

1	2	3	4	5
male			female	

61. Mrs. Stanley will be X years old 5 years hence. How old was she 5 years ago?

1	2	3	4	5
male			female	

62. A cook has 28 pounds of sugar which she must separate into packages containing 14 ounces each. How many such packages can she make?

1	2	3	4	5
male			female	

63. To raise $500. for an orphanage, the Girl Scouts plan a musical festival with expenses of $250. What is the minimum number of tickets at 75 cents each that will have to be sold to reach their goal?

1	2	3	4	5
male			female	

64. If one pie serves 7 women at a luncheon, how many pies are needed to serve a luncheon of 91 women?

1	2	3	4	5
male			female	

65. The city of Cleveland has 154 Brownie troops taking part in a cookie sale. On a certain day, 40 troops have met their sales quota and 20 have sold no cookies at all. How many of the remaining troops must meet their quota so that approximately 65% of the troops have done so?

1	2	3	4	5
male			female	

66. The indicator of an oil tank shows 1/5 full. After the truck delivers 165 gallons of oil the indicator shows 4/5 full. What is the capacity of the tank (in gallons)?

1	2	3	4	5
male			female	

67. To send a parcel to Zone 7 the cost is 30¢ for the first pound and 15.5¢ for each additional pound. What is the cost of sending a package weighing 48 ounces to Zone 7?

1	2	3	4	5
male			female	

68. A broad jumper makes an average standing jump of 8 feet. In how many jumps will he cover y yards.

1 2 3 4 5
male female

69. Three times a number less seven is thirty-two. What is twice the number?

1 2 3 4 5
male female

70. If a pipe fills a tank in h hours, what part of the tank does it fill in 2 hours?

1 2 3 4 5
male female

71. A man uses 1/2 of his available oil to heat his home during 1 week of extreme cold weather. If the tank was 3/4 full at the beginning of the week, what part of the full capacity did he use this week?

1 2 3 4 5
male female

72. If the registration of a school increased from 300 to 1200, what is the percent of increase in registration?

1 2 3 4 5
male female

73. A boy makes a diagram of a plane to a scale of 1/400 of actual size. If the diagram of the plane is 9.6 inches in length, what is the length, in feet, of the plane?

1 2 3 4 5
male female

74. If two cups of melted chocolate weigh 16 ounces, how many cups of melted chocolate can be obtained from a package weighing 1 pound 8 ounces?

1 2 3 4 5
male female

75. If the average ages of three people is 44 years, and if no one of them is less than 42 years old, what is the maximum age (in years) of any one person?

1 2 3 4 5
male female

76. A woman works d days and earns w dollars more than p dollars. What are her average earnings per day?

1 2 3 4 5
male female

77. The water which is 7 inches high in a fish tank 1-1/4 feet by 8 inches, is poured into a tank 13 inches by 20 inches. What height will it reach in the larger tank?

1 2 3 4 5
male female

78. One light flashes every ten minutes and a second light flashes every six minutes. If both lights flash at 5:00 PM, what is the next time they both flash?

1 2 3 4 5
male female

79. The secretary of a club can address the envelopes for a mailing in 40 minutes. Her younger sister who could do the entire job alone in 1 hour assists her. How long will it take to complete the job if both girls work?

1 2 3 4 5
male female

80. If the taxi fare is c cents for the first quarter mile and s cents for each additional quarter mile, what is the charge for a trip of X miles (where X is greater than 1)?

1 2 3 4 5
male female

81. Three people invest $2000., $3000., and $5000. respectively upon the formation of a business partnership. The net profits at the end of the first year amount to $960. How much should the person who invested the least money receive as his fair share if the profits are divided in accordance with the amount each person invested?

1	2	3	4	5
male			female	

82. A major league baseball team plays 154 games in a season. On a certain date, a team has won 40 and lost 20 games. How many of the remaining games must they win to finish the season winning approximately 65% of their games?

1	2	3	4	5
male			female	

83. The center of a basketball team is 6'3". The two guards are each 6'. One forward is 5'10" and the other is 5'11". What is the average height of the basketball team?

1	2	3	4	5
male			female	

84. A girl keeps a record of time she spends practicing her music. On Monday she spends 1-1/4 hours; on Tuesday and Wednesday she spend 2 hours each day; on Thursday she spends 1-3/4 hours. How many hours will she have to spend practicing for the rest of the week in order to have an average of exactly 90 minutes each day?

1	2	3	4	5
male			female	

85. Because of a decrease in available man power, a tool factory reduced its monthly output by 20%. What was the necessary percent increase of man power needed to bring the output up to normal production?

1	2	3	4	5
male			female	

86. For a 9 week period a busboy working 5 days per week earned $454.50. What were his earnings per day?

1	2	3	4	5
male			female	

87. If bridal shower invitations are sold at the rate of 3 for 10¢, what will be the cost of 3 dozen such invitations?

1	2	3	4	5
male			female	

88. If a woman was r years old s years ago, how many years old will she be t years from now?

1	2	3	4	5
male			female	

89. Three boys have marbles in the ratio of 19:5:3. If the boy with the least number has 9 marbles, how many marbles does the boy with the greatest number have?

1	2	3	4	5
male			female	

90. If crocheted scarves bought at prices ranging from $2.00 to $3.50 are sold in a boutique at prices ranging from $3.00 to $4.25, what is the greatest possible profit that might be made in selling 8 scarves?

1	2	3	4	5
male			female	

91. A man owned 5/8 of an interest in a house. He sold 1/5 of his interest, at cost, for $1000. What is the total value of the house?

1	2	3	4	5
male			female	

92. When the gasoline gauge of an automobile shows 1/8 full it takes 14 gallons to completely fill the tank. What is the capacity of the gasoline tank (in gallons)?

1	2	3	4	5
male			female	

93. Susan cuts 2/3 of a yard of licorice into 12 strips. What part of a foot would two strips be?

1	2	3	4	5
male			female	

94. A baseball team won W games and lost L games. What fractional part of its games did it win?

1	2	3	4	5
male			female	

95. If x must be greater than 4, which of the following must have the least value?

1	2	3	4	5
male			female	

96. Suppose z- 3x/y, what happens to z if x is doubled and y is tripled?

1	2	3	4	5
male			female	

97. Judy gave her sister Lucy $12.00 which is 15% of the amount she earned babysitting. How much does Judy have now?

1	2	3	4	5
male			female	

98. At the end of the Campfire Girls annual candy sale, candy which was originally sold for $1.60 per pound is offered for 48 cents in a six ounce package. What is the ratio of the former pricc to the present price?

1	2	3	4	5
male			female	

99. 1 Angstrom unit = .0001 micron. How many Angstrom units are there in .01 micron?

1	2	3	4	5
male			female	

100. A man deposits $700.00 in a bank which pays 3% interest per year. How much money does he have at the end of the year?

1 2 3 4 5
male female

101. In the dining room of a children's camp, a quart of milk can serve 4 campers or 3 counselors. If at one meal 40 campers and 12 counselors are being served, how much milk will remain unused if 16 quarts are brought into the dining room?

1 2 3 4 5
male female

102. A boy has 85 cents in nickels and dimes, 12 coins in all; how many coins are nickels?

1 2 3 4 5
male female

103. Mr. Goodsale receives a salary of $6000. per year plus 5% of all his sales over $10,000. and a special bonus of $500. if his sales exceed $20,000. What are his earnings during a year when his sales total $21,000?

1 2 3 4 5
male female

104. A group of Boy Scouts hiked into the woods at a rate of 4 miles an hour, and returned over the same road at the rate of 3 miles an hour. If they completed the entire hike in 3-1/2 hours, how far (in miles) into the woods did they hike?

1 2 3 4 5
male female

105. A boy rides his bicycle ten miles at an average rate of ten miles an hour. The average rate for the entire trip is approximately?

1 2 3 4 5
male female

106. 7 lbs. of pears cost as much as 10 lbs. of apples and one pound of oranges. 7 lbs. of oranges cost as much as one pound of pears and 2 lbs. of apples. How many lbs. of apples may be purchased for the amount of money required to purchase 12 lbs. of pears?

1 2 3 4 5
male female

107. The Best Co. linen department formerly sold washcloths for 80¢ each, but now offers them at $9.00 per dozen. What is the ratio of the old price to the new price?

1 2 3 4 5
male female

108. What fraction must be subtracted from the sum of 1/2 and 1/3 to have an average of 1/6?

1 2 3 4 5
male female

109. What part of a quarter is two pennies, two nickels and one dime?

1 2 3 4 5
male female

110. A student's marks are 70, 90, 65, 85, and 75. What must his mark be on the next test to raise his average to 80%.

1 2 3 4 5
male female

111. The blueprint of a room is drawn to the scale 1 inch equals 20 feet. If a room is actually 10 yards long, how long is the line of the blueprint drawn to represent the lenght of the room?

1 2 3 4 5
male female

112. What is the maximum total weight of ten eggs, if four of them weigh 15 to 25 ounces each and the others weigh from 20 to 25 ounces each?

1	2	3	4	5
male			female	

113. Barbara spends $11.22 purchasing and mailing party invitations. If the invitations are purchased for 3 for 10¢, and each requires an 8¢ stamp, how many invitations were purchased?

1	2	3	4	5
male			female	

114. A department store offers a vacuum cleaner for $72.00 instead of the list price of $90.00. What is the rate of discount for this item?

1	2	3	4	5
male			female	

115. How many 8 cent stamps may be purchased for c cents?

1	2	3	4	5
male			female	

116. A trackman runs 100 yards in 9.8 seconds and then runs 440 yards in 49 seconds. What is the ratio of his average speed in the 100 yard run?

1	2	3	4	5
male			female	

117. At a Girl Scout masquerade party the troop leaders judge the girls costumes. If the judges eliminate 1/4 of the eligible contestants after each half hour, and 256 girls were wearing costumes, how many would still be eligible for a prize after two hours?

1	2	3	4	5
male			female	

118. Carol is running for class secretary. If there are 356 students who vote to choose the secretary, and a total of 5 candidates seeking the office, what is the least number of votes Carol must receive and still have more votes than any other candidate?

1	2	3	4	5
male			female	

119. An electric saw sells for $65.00. This price gives the hardware store owner a profit of 30% on his costs. What will be the new retail price if he cuts his profits to 10% of costs?

1	2	3	4	5
male			female	

120. ?/32 = .625

1	2	3	4	5
male			female	

APPENDIX II

The procedure used for designation of item content as male or female was developed by the author in a Pilot Study conducted at Metuchen High School, Metuchen, New Jersey. The pilot was conducted for a dual purpose: First to ascertain the general item content seen as male or female by a sample of the population in question, and second to discover if male and female subjects viewed the content of the same item differently. That is, would male subjects view a given item as more male (or female) than female subjects viewed the same item?

The subjects were presented with 30 mathematics problem solving items (which follow) which they were asked to judge for content familiarity to males or females using the procedure described in Stage 3. Mean judgement scores were computed for the male and female subgroups separately for each item. These scores are reported in Table 1A.

TABLE 1A

Mean Judgement Scores

Item	Male Mean	Female Mean
1	2.2	2.2
2	2.1	2.1
3	4.4	4.7
4	4.1	4.4
5	3.5	3.5
6	1.9	1.4
7	4.1	4.5
8	3.4	3.1
9	2.7	2.6
10	2.6	2.5
11	3.2	2.5
12	3.6	3.1
13	3.1	2.9
14	2.1	1.7
15	2.8	2.5
16	4.4	4.6
17	4.1	4.3
18	3.4	3.6
19	2.6	2.7
20	2.0	1.3
21	4.2	4.7
22	3.4	3.6
23	2.7	2.3
24	3.7	4.0
25	2.0	1.7
26	2.5	2.1
27	3.3	3.5
28	4.3	4.7
29	3.4	3.3
30	1.9	1.8

DIRECTIONS: Read each of the following items: As you read them do not focus your attention on the mathematics involved, but rather, on the content of the item. After reading the item decide whether you believe that the content would, in general, be more familiar to males or to females. Judge each item on the sliding scale provided, where 1 would indicate content very familiar to males, 3 would indicate content which is neither more familiar to males or females, and 5 would indicate content very familiar to females. (The males and females for whom you are judging would be approximately your age.)

1. A mortorist leaves at 9:00 A.M. and stops for repairs at 9:20 A.M. If the distance covered was 18 miles, what was the average velocity for his part of the trip?

   1 2 3 4 5
   male female

2. If a pipe fills a tank in h hours, what part of the tank does it fill in 2 hours?

   1 2 3 4 5
   male female

3. A secretary types business letters at a rate of d words per minute, and statistical tables at a rate of u words per minute. What is her average rate of typing if she types for 2 hours?

   1 2 3 4 5
   male female

4. In planning the annual Girl Scout Jamboree, Mrs. Smith donates 15 pounds of hot dogs, and Mrs. Jones donates 9 pounds. Together they represent 16-2/3% of the total number of pounds needed. How many more pounds must be donated to assure the organizers that at least .750 of the amount needed is available?

   1 2 3 4 5
   male female

5. Mrs. Stanley will be X years old 5 years hence. How old was she 5 years ago?

   1 2 3 4 5
   male female

6. An architect uses a scale of 1/4 inch to a foot. If the dimensions of a room as represented on the plans of a house are 4-1/4 inches by 3-3/4 inches, the actual dimensions of the room in feet are what?

1 2 3 4 5
male female

7. The batter which is 7 inches high in a loaf pan one and one-quarter feet by 8 inches, is poured into a pan 13" by 20". What height will it reach in the larger pan?

1 2 3 4 5
male female

8. Three women invest $2,000.00, $3,000.00, and $5,000.00 respectively upon the formation of a partnership in a boutique. The net profits at the end of the first year amount to $960.00 How much should the women who invested the least money receive as her share if the profits are divided in accordance with the amount each woman invested?

1 2 3 4 5
male female

9. A merchant paid $30.00 for an article. He wishes to place a price tag on it so that he can offer a 10% discount on the price marked on the tag and still make a profit of 20% on the cost. What price should he mark on the tag?

1 2 3 4 5
male female

10. A man binds 112 books in one day, and his assistant works one quarter as fast. How many days would it take them to bind 560 books if each works alone on alternate days?

1 2 3 4 5
male female

11. A corporation has 8 departments each with 10-16 bureaus. In each bureau there are at least 40, but no more than 60 workers. If 10% of the workers in each bureau are typists, what is the minimum number of typists in a department?

1 2 3 4 5
male female

12. A woman works d days and earns w dollars more than p dollars. What are her average earnings per day?

1 2 3 4 5
male female

13. If Mrs. Anderson leaves her home at point R and drives 14 miles directly north to S, then 8 miles directly east to T, what is the straight line distance from T to R in miles?

1 2 3 4 5
male female

14. A car uses a gallon of gasoline in traveling 15 miles. Another automobile can travel m miles on a gallon of gasoline. How many miles can the second travel on the amount of gasoline required by the first car in going 60 miles?

1 2 3 4 5
male female

15. One-tenth of a man's salary is spent for clothing, one-third for food, and one-fifth for rent. What percent of the salary is left for other expenditures and savings?

1 2 3 4 5
male female

16. In making drapes Joan needs to cut 8 foot panels. How many panels may she cut from fabric that is y yards long?

1 2 3 4 5
male female

17. The city of Cleveland has 154 Brownie troops taking part in a city-wide cookie sale. On a certain date 40 troops have met their sales quota and 20 have sold no cookies at all. How many of the remaining troops must meet their quota so that approximately 65% of the troops have done so?

1 2 3 4 5
male female

18. If a woman was r years old s years ago, how many years old will she be t years from now?

1 2 3 4 5
male female

19. If the taxi fare is c cents for the first quarter of a mile and s cents for each additional quarter of a mile, what is the charge (in cents) for a trip of X miles (where X is greater than 1)?

1 2 3 4 5
male female

20. The center of a basketball team is 6'3". The two guards are each 6'. One forward is 5'10" and the other is 5'11". What is the average height of the basketball team?

1 2 3 4 5
male female

21. If 2s seamstresses can sew 2d dresses in 2w weeks, how many seamstresses would it take to sew 4d dresses in 4w weeks?

1 2 3 4 5
male female

22. If the average of the ages of three women is 44 years, and if no one of them is less than 42 years old, what is the maximum age (in years) of any one women?

1 2 3 4 5
male female

23. Mr. A owes Mr. B $70. and Mr. B. owes Mr. A. $60. If Mr. A. gives Mr. B. a $50. bill how many dollars in change should Mr. B. give Mr. A?

1 2 3 4 5
male female

24. Barbara and Arlene volunteer to bring equal amounts of soda to the senior class party. Arlene has bought twelve four-quart bottles. If Barbara gets her soda in three-quart bottles how many must she buy?

1 2 3 4 5
male female

25. A carpenter is supposed to build a minimum of t tables. In d days he builds m more than the minimum. What is the average number of tables he produced each day?

1	2	3	4	5
male			female	

26. An automobile passes City X at 9:55 A.M. and City Y at 10:15 A.M. City X is 30 miles from City Y. What is the average rate of the auto in miles per hour?

1	2	3	4	5
male			female	

27. Judy reads at the rate of 55 pages per hour. How many minutes will it take her to read 2/5 of a page at this rate?

1	2	3	4	5
male			female	

28. A woman can baste a 10 inch seam by hand in 5 minutes, and go over it on a machine in 2 minutes. What is her average rate in inches per minute to finish the seam?

1	2	3	4	5
male			female	

29. There are 45 pupils in a certain English class. If two-thirds of the pupils are girls, and one half of the girls are blue-eyed, how may blue-eyed girls are in the class?

1	2	3	4	5
male			female	

30. 7 type p nails are worth 5 type q nails. 3 type p nails are worth x type q nails. What is the numerical value of x?

1	2	3	4	5
male			female	

Judgement score means were translated into ratings using the criteria described previously. Items with means in the 1.0-2.0 range were designated as male, 2.5-3.5 as neutral, and 4.0-5.0 as female. All other items were categorized as ambiguous. Table 2A shows the ratings earned by each item from the male and female subgroups.

TABLE 2A

Mean Judgement Scores

Item	Male	Female
1	A	A
2	A	A
3	F	F
4	F	F
5	N	N
6	M	M
7	F	F
8	N	N
9	N	N
10	N	N
11	N	N
12	A	N
13	N	N
14	A	M
15	N	N
16	F	F
17	F	F
18	N	A
19	N	N
20	M	M
21	F	F
22	N	A
23	N	A
24	A	F
25	M	M
26	N	A
27	N	N
28	F	F
29	N	N
30	M	M

An anlaysis of the results indicates that two items were seen as ambiguous (A) by both sex subgroups, but seven other items were seen as ambiguous by one subgroup or the other. Ten items were rated as neutral (N) by both groups, with 5 other items rated as neutral by one of the subgroups. Seven items were rated as female (F) by both subgroups, and four were seen as male (M). In only seven out of the 30 cases did the two sex subgroups vary sufficiently in their mean judgement scores to cause the judgement ratings to differ. It seems reasonable, therefore, to assume that, in general, male and female subjects view the familiarity of content to the sexes similarly. Based on this outcome, the decision was made to eliminate any item viewed (rated) differently by the two sex subgroups.

An analysis of the content viewed as male as opposed to female or neutral indicated that items in which only noun or pronoun referrents were sex-linked were viewed as neutral. The item:

> "Mrs. Stanley will be X years old 5 years hence.
> How old was she 5 years ago?"

is an item of this type. Items dealing with money were also viewed as neutral.

Female items dealt with traditional female skills such as typing, organizing picnics, cooking and sewing. Male items included content dealing with architect's scales, sports, and building. As a result of these findings, the construction of items described in Stage 1 was undertaken using sterotyped sex-linked pursuits as content for potentially male and female items.

APPENDIX III

Math Attitude Scale

Lewis R. Aiken, Jr.

Women's College of the University of North Carolina

Greensboro, North Carolina

and Ralph Mason Dreger

Jacksonville University

Jacksonville, Florida

NAME ________________________

O P I N I O N N A I R E

DIRECTIONS: Please write your name in the upper right hand corner. Each of the statements on this opinionnaire expresses a feeling which a particular person has toward mathematics. You are to express, on a five-point scale, the extent of agreement between the feeling expressed in each statement and your own personal feeling. The five points are: Strongly disagree (SD), Disagree (D), Undecided (U), Agree (A), Strongly agree (SA). You are to encircle the letters(s) which best indicates how closely you agree or disagree with the feeling expressed in each statement AS IT CONCERNS YOU.

1. I do not like mathematics. I am always under a terrible strain in a math class.	SD	D	U	A	SA
2. I do not like mathematics, and it scares me to have to take it.	SD	D	U	A	SA
3. Mathematics is very interesting to me. I enjoy math courses.	SD	D	U	A	SA
4. Mathematics is fascinating and fun.	SD	D	U	A	SA

5. Mathematics makes me feel secure, and at the same time it is stimulating.	SD	D	U	A	SA
6. I do not like mathematics. My mind goes blank, and I am unable to think clearly when working math.	SD	D	U	A	SA
7. I feel a sense of insecurity when attempting mathematics.	SD	D	U	A	SA
8. Mathematics makes me feel uncomfortable, restless, irritable, and impatient.	SD	D	U	A	SA
9. The feeling that I have toward mathematics is a good feeling.	SD	D	U	A	SA
10. Mathematics makes me feel as though I'm lost in a jungle of numbers and can't find my way out.	SD	D	U	A	SA
11. Mathematics is something which I enjoy a great deal.	SD	D	U	A	SA
12. When I hear the word math, I have a feeling of dislike.	SD	D	U	A	SA
13. I approach math with a feeling of hesitation--hesitation resulting from a fear of not being able to do math.	SD	D	U	A	SA
14. I really like mathematics.	SD	D	U	A	SA
15. Mathematics is a course in school which I have always liked and enjoyed studying.	SD	D	U	A	SA
16. I don't like mathematics. It makes me nervous to even think about having to do a math problem.	SD	D	U	A	SA
17. I have never liked math, and it is my most dreaded subject.	SD	D	U	A	SA
18. I love mathematics. I am happier in a math class than in any other class.	SD	D	U	A	SA

19. I feel at ease in mathematics, and I like it very much.	SD	D	U	A	SA
20. I feel a definite positive reaction to mathematics; it's enjoyable.	SD	D	U	A	SA

APPENDIX IV

Name ______________________________ Sex (Circle one) Male Female
Grade (Circle one) 10th 11th 12th Birth date ____________
Math Teachers Name ________________________
Do you plan to go to college? (Circle one) Yes No Undecided
What is your favorite subject in school? ____________________
In what subject do you get your best grades? ________________

TEST OF MATHEMATICS APTITUDE

The test you are about to take is a measure of mathematics aptitude, very similar in form to the College Boards Scholastic Aptitude Test. Your scores on this test will be returned to you, and to your mathematics teacher. For you, the test can serve as practice for the College Boards (if you have not already taken them), and as an indication of your ability in mathematics. For your teacher, the results will aid him/her in developing a better understanding of your strengths and weaknesses in mathematics. You are urged to take the test seriously, and do your best.

You will have the full period for this test. You may use scratch paper, which will be provided by the teacher. If you are not sure how to do any problem, go on to the next one (they do not get progressively more difficult). Each item has 5 choices. Place the letter corresponding to the correct answer on the line provided for it next to the item.

1. A fishing rod manufacturer finds that 0.4% of his production is defective and not suitable for marketing. How many rods of each 1000 produced will be rejected?
(a) 4 (b) 14 (c) 40 (d) 140 (e) 400 1. ________

2. A recipe for cheesecake calls for 12 ounces of cottage cheese and 6 ounces of cream cheese. If there are 216 calories in 8 ounces of cottage cheese, and 106 calories in 1 ounce of cream cheese, how many calories are there in 36 ounces of the cheese mixture?
(a) 960 (b) 972 (c) 1920 (d) 3816 (e) 4788 2. ________

3. $r = 18$
$\frac{r}{s} = \frac{6}{y}$
$\frac{y}{s} = ?$
(a) $\frac{1}{3}$ (b) $3\frac{1}{6}$ (c) $3\frac{1}{3}$ (d) 6 (e) 18 3. ________

4. At a Girl Scout masquerade party the troop leaders judge the girls costumes. If the judges eliminate $\frac{1}{4}$ of the eligible contestants after each half hour, and 256 girls were wearing costumes, how many would still be eligible for a prize after two hours?
(a) 0 (b) 16 (c) 32 (d) 64 (e) 81 4. ________

5. What fraction must be subtracted from the sum of $\frac{1}{2}$ and $\frac{1}{3}$ to have an average of $\frac{1}{6}$?
(a) $\frac{1}{3}$ (b) $\frac{2}{5}$ (c) $\frac{1}{6}$ (d) $\frac{5}{6}$ (e) 1 — 5. ________

6. How many 8 cent stamps may be purchased for C cents?
(a) $\frac{8}{C}$ (b) C (c) 8C (d) $\frac{C}{8}$ (e) C-8 — 6. ________

7. If two halves of 2½ is added to 2½ the result is?
(a) 2½ (b) 3 (c) 5 (d) 5½ (e) 6 — 7. ________

8. A cook has 28 pounds of sugar which she must separate into packages containing 14 ounces each. How many such packages can she make?
(a) 2 (b) 4 (c) 8 (d) 16 (e) 32 — 8. ________

9. A corporation has 8 departments with 10-16 bureaus each. In each bureau there are at least 40, but no more than 60 workers. If 10% of the workers in each bureau are typists, what is the minimum number of typists in a department?
(a) 40 (b) 65 (c) 96 (d) 320 (e) 768 — 9. ________

10. Mrs. Crocker can do the dinner dishes in a quarter of an hour, while her daughter takes half an hour. What part of an hour would it take to do the dishes if they work together?
(a) $\frac{1}{8}$ (b) $\frac{1}{6}$ (c) $\frac{3}{16}$ (d) $\frac{3}{8}$ (e) $\frac{4}{3}$ — 10. ________

11. Ten merchants agreed to purchase uniforms for a local baseball team at a total cost of m dollars. After a disagreement, 2 merchants dropped out of the project. By how many dollars was the cost to each remaining merchant increased?
(a) $\frac{m-20}{2}$ (b) $\frac{m}{40}$ (c) $\frac{m}{9}$ (d) $\frac{m}{2}$ (e) 2m — 11. ________

12. If crocheted scarves bought at prices ranging from \$2.00 to \$3.50 are sold in a boutique at prices ranging from \$3.00 to \$4.25, what is the greatest possible profit that might be made in selling 8 scarves?
(a) \$2.50 (b) \$4.00 (c) \$6.00 (d) \$9.00 (e) \$18.00 — 12. ________

13. A baseball team won W games and lost L games. What fractional part of its games did it win?
(a) $\frac{L}{W}$ (b) $\frac{W-L}{W}$ (c) $\frac{W}{L}$ (d) $\frac{W+L}{W}$ (e) $\frac{W}{W+L}$ 13. ________

14. Sue bought 8 feet of ribbon to use to decorate baskets. How many ½ inch strips will she be able to cut?
(a) 24 (b) 32 (c) 200 (d) 320 (e) 384 14. ________

15. A motorist leaves at 9:00 AM and stops at 9:20 AM. If he covered 18 miles, what was his average velocity for this part of the trip?
(a) 5.4 (b) 6 (c) 54 (d) 36 (e) 60 15. ________

16. After several tryouts, 20% of a football squad was discharged. The coach then had 32 players. How many players were on the squad at first?
(a) 24 (b) 26 (c) 39 (d) 40 (e) 80 16. ________

17. A picture in an art museum is six feet wide and eight feet long. If its frame has a width of six inches, what is the ratio of the area of the frame to the area of the picture?
(a) $\frac{5}{16}$ (b) $\frac{5}{4}$ (c) $\frac{4}{5}$ (d) $\frac{5}{12}$ (e) $\frac{3\frac{1}{5}}{1}$ 17. ________

18. A major league baseball team plays 154 games in a season. On a certain date, a team has won 40 and lost 20 games. How many of the remaining games must they win to finish the season winning approximately 65% of their games?
(a) 60 (b) 80 (c) 90 (d) 94 (e) 100 18. ________

19. A trackman runs 100 yards in 9.8 seconds and then runs 440 yards in 49 seconds, What is the ratio of his average speed in the 100 yard run?
(a) $\frac{49}{43}$ (b) $\frac{22}{25}$ (c) $\frac{11}{12}$ (d) $\frac{12}{11}$ (e) $\frac{25}{22}$ 19. ________

20. During the month of June, the General Motors Corporation sold twice as many Chevelles as Impalas, 3 times as many Vegas as Chevelles, and 4 times as many Novas as Vegas. If they sold 300 more Vegas than Chevelles, how many Novas were sold?
(a) 75 (b) 180 (c) 300 (d) 720 (e) 1800 20. ________

21. If one pie serves 7 women at a luncheon, how many pies are needed to serve a luncheon of 91 women?
(a) 7 (b) 9 (c) 13 (d) 15 (e) 657 21. ________

22. In planning for the church picnic, the ladies guild purchases enough hot dogs to allow each of the 18 persons 3 ho t dogs. If 9 additional perscns come to the picnic, how many fewer hot dogs will each person be given?
(a) 1 (b) $1\frac{1}{2}$ (c) 2 (d) $2\frac{1}{4}$ (e) $2\frac{1}{2}$ 22. ________

23. A car uses a gallon of gasoline in traveling 15 miles. Another automobile can travel m miles on a gallon of gasoline. How many miles can the second travel on the amount of gasoline required by the first car in going 60 miles?
(a) $\frac{m}{4}$ (b) m (c) 4m (d) $\frac{m}{9}$ (e) 9m 23. ________

24. A mile runner is clocked at 58 seconds after the first quarter mile (440 yards), and 1 minute 56 seconds after he has run a half mile. What rate (in yards per second) must he maintain for the final two quarters if he is to run a 4 minute mile?
(a) 7.0 (b) 7.1 (c) 7.3 (d) 7.6 (e) 7.8 24. ________

25. The Japanese ken is equivalent to 5.97 feet. How many feet are there in 59.7 ken?
(a) 0.1 (b) 10 (c) 248 (d) 356 (e) 360 25. ________

26. An architect uses a scale of $\frac{1}{4}$ inch to a foot. If the dimensions of a room as represented on the plans of a house are $4\frac{1}{4}$ inches by $3\frac{3}{4}$ inches, the actual dimensions of the room in feet are?
(a) $8\frac{1}{2}$x$9\frac{1}{2}$ (b) 17x15 (c) 34x30 (d) 51x45 (e) 68x60 26. ________

27. Suppose $z = \frac{3x}{y}$, what happens to z if x is doubled and y is tripled?
(a) doubled (b) tripled (c) halved (e) multiplied by 6
(e) multiplied by a factor of $\frac{2}{3}$
27. ________

28. A woman works d days and earns w dollars more than p dollars. What are her average earnings per day?
(a) $\frac{p}{w+d}$ (b) $\frac{d+p}{w}$ (c) $\frac{d+w}{p}$ (d) $\frac{p+d}{w}$ (e) $\frac{p+w}{d}$ 28. ________

29. The Best Co. linen department formerly sold washclothes for 80¢ each, but now offers them at $9.00 per dozen. What is the ratio of the old price to the new price?
(a) 5:1 (b) 16:15 (c) 15:16 (d) 8:45 (e) 45:8 29. ________

30. $\frac{?}{32} = .625$
(a) 16 (b) 20 (c) 22 (d) 24 (e) 26 30. ________

31. Jane cut a length of cloth 13 yards, 5 feet and 1 inch into three equal parts. The length of each piece is?
(a) 4 yards 4 feet (b) 4 yards $4\frac{1}{3}$ feet
(c) 4 yards 2 feet $8\frac{1}{3}$ inches
(d) 4 yards 1 foot (e) 4 yards $5\frac{1}{3}$ feet 31. ________

32. A student reads at the rate of 50 pages per hour. How many minutes will it take him to read $\frac{2}{5}$ of a page at this rate?
(a) 0.2 (b) .48 (c) 2.2 (d) 13.5 (e) 22 32. ________

33. If 2 cups of melted chocolate weigh 16 ounces, how many cups of melted chocolate can be obtained from a package weighing 1 pound 8 ounces?
(a) 2.2 (b) 3 (c) 3.2 (d) 4 (e) 4.3 33. ________

34. The indicator of an oil tank shows $\frac{1}{5}$ full.
After the truck delivers 165 gallons of oil the indicator shows $\frac{4}{5}$ full. What is the capacity of the tank (in gallons)?
(a) 55 (b) 105 (c) 140 (d) 175 (e) 275 34. ________

35. There are 45 pupils in a certain class. If two-thirds of the pupils are boys , and one half of the boys have blue eyes, how many blue eyed boys are in the class?
(a) 15 (b) 30 (c) 34 (d) 38 (e) 43 35. ________

36. A secretary types business letters at a rate of d words per minute, and statistical tables at a rate of u words per minute. What is her average rate of typing if she types for 2 hours?
(a) $\frac{du}{2}$ (b) $\frac{d+u}{2}$ (c) $\frac{du}{d+2}$ (d) $\frac{2du}{d+u}$ (e) $\frac{d+u}{2du}$ 36. ________

37. If bridal shower invitations are sold at the rate of 3 for 10¢, what will the cost be for 3 dozen such invitations?
(a) $.42 (b) $1.08 (c) $1.20 (d) $1.26
(e) $1.44 37. ________

38. A carpenter is supposed to build a minimum of t tables. In d days he builds m more tables than the minimum. What is the average number of tables he produced each day?
(a) $\frac{m}{d}$ (b) $\frac{t}{d}$ (c) $\frac{m+t}{d}$ (d) $\frac{t-m}{d}$ (e) $\frac{m-t}{d}$ 38. ________

39. What is the maximum total weight of ten eggs, if four of them weigh 15 to 25 ounces each and the others weigh from 20 to 25 ounces each?
(a) 180 oz. (b) 210 oz. (c) 220 oz.
(d) 225 oz. (e) 250 oz. 39. ________

Name ______________________________ Sex (Circle one) Male Female
Grade (Circle one) 10th 11th 12th Birth date ______________
Math Teachers Name ______________________________

This is the final part of the mathematical aptitude test which you began yesterday. You will have the full period for this test. You may use scratch paper, which will be provided by the teacher. If you are not sure how to do any problem, go on to the next one (they do not get progressively more difficult). Each item has 5 choices. Place the letter corresponding to the correct answer on the line provided for it next to the item.

1. A group of Boy Scouts hiked into the woods at a rate of 4 miles an hour, and returned over the same road at the rate of 3 miles an hour. If they completed the entire hike in $3\frac{1}{2}$ hours, how far (in miles) into the woods did they hike?
(a) 3 (b) 4 (c) 5 (d) 6 (e) 7 1. ________

2. In planning the annual Girl Scout Jamboree, Mrs. Smith donates 15 pounds of hot dogs, and Mrs. Jones donates 9 pounds. Together they represent $16\frac{2}{3}$ % of the total number of pounds needed. How many more pounds must be donated to assure the organizers that at least .75 of the amount needed is available?
(a) 28 (b) 75 (c) 80 (d) 87 (e) 84 2. ________

3. If two parts of sand are mixed with three parts of gravel, what part of the total mixture is sand?
(a) $\frac{1}{3}$ (b) $\frac{2}{5}$ (c) $\frac{3}{5}$ (d) $\frac{2}{3}$ (e) $\frac{3}{2}$ 3. ________

4. Susan cuts $\frac{2}{3}$ of a yard of licorice into 12 strips. What part of a foot would two strips be?
(a) $\frac{1}{9}$ (b) $\frac{1}{6}$ (c) $\frac{1}{3}$ (d) $\frac{1}{2}$ (e) $\frac{2}{3}$ 4. ________

5. In a library system having six branches there are 60 workers employed. If no library has less than 7 workers and no more than 18, what is the minimum number of workers in any two of these branches?
(a) 10 (b) 14 (c) 20 (d) 25 (e) 36 5. ________

6. If a woman was r years old s years ago, how many years old will she be t years from now?
(a) rs+t (b) r-s+t (c) r+s+t (d) rst (e) s-r+t 6. ________

7. If x must be greater than 4, which of the following must have the least value?
(a) $\frac{4}{x+1}$ (b) $\frac{4}{x-1}$ (c) $\frac{4}{x}$ (d) $\frac{x}{4}$ (e) $\frac{x+1}{4}$ 7. ________

8. At the end of the Campfire Girls annual candy sale, candy which was originally sold for $1.60 per pound is offered for 48 cents in a six ounce package. What is the ratio of the former price to the present price?
(a) 3:1 (b) 5:4 (c) 10:3 (d) 3:10 (e) 4:5 8. ________

9. Ten broken hens eggs weigh one pound. What is the weight in ounces of 1 broken hens eggs?
(a) 0.1 (b) 1.6 (c) 1.8 (d) 10 (e) 16 9. ________

10. A department store had an end of winter sale on ladies coats at the following prices:
A- $80. coats reduced to $55.
B- $85. coats reduced to $60.
C- $90. coats reduced to $65.
D- $95. coats reduced to $70.
E- $120. coats reduced to $95.
Which group of coats was offered at the greatest rate of discount from the original price?
(a) A (b) B (c) C (d) D (e) E 10. ________

11. A boy rides his bicycle ten miles at an average rate of ten miles an hour. The average rate for the entire trip is approximately?
(a) 2 (b) 10 (c) 11 (d) 21 (e) 22 11. ________

12. Barbara spends $11.22 purchasing and mailing party invitations. If the invitations are purchased for 3 for 10¢, and each requires an 8¢ stamp, how many invitations were purchased?
(a) 11 (b) 22 (c) 33 (d) 66 (e) 99 12. ________

13. The center of a basketball team is 6'3". The two guards are each 6'. One forward is 5'10", and the other is 5'11". What is the average height of the basketball team?
(a) 5.2' (b) 5.56' (c) 5.6' (d) 6' (e) 6.2' 13. ________

14. In making tuna fish salad, a recipe calls for 2 cups of fish to $\frac{1}{8}$ of a cup of chopped celery. How much celery should be used for 8 cups of tuna fish?
(a) $\frac{1}{8}$ (b) $\frac{1}{4}$ (c) $\frac{1}{2}$ (d) 2 (e) 4 — 14. ________

15. A broad jumper makes an average standing jump of 8 feet. In how many jumps will he cover y yards?
(a) $\frac{y}{8}$ (b) $\frac{3y}{8}$ (c) 8y (d) $\frac{8y}{3}$ (e) 24y — 15. ________

16. Because of a decrease in available man-power, a tool factory reduced its monthly output by 20%. What was the necessary percent increase of man-power needed to bring the output up to normal production?
(a) 20 (b) 25 (c) 50 (d) 120 (e) 125 — 16. ________

17. What part of a quarter is two pennies, two nickels, and one dime?
(a) $\frac{3}{25}$ (b) $\frac{22}{25}$ (c) $\frac{1}{22}$ (d) $\frac{1}{5}$ (e) $\frac{3}{5}$ — 17. ________

18. If 8 men can build a house in 12 days, what is the percentage increase in the number of days required to do the job if 2 men are laid off?
(a) 25 (b) 33 (c) 50 (d) 67 (e) 75 — 18. ________

19. The blueprint of a room is drawn to the scale 1 inch equals 20 feet. If a room is actually 10 yards long, how long is the line of the blueprint drawn to represent the length of the room?
(a) $\frac{2}{3}$ inch (b) $1\frac{1}{2}$ inch (c) $1\frac{1}{3}$ inch (d) 2 inches
(e) 6 inches — 19. ________

20. Mr. Goodsale receives a salary of $6000. per year plus 5% of all his sales over $10,000., and a special bonus of $500. if his sales exceed $20,000. What are his earnings during a year when his sales total $21,000?
(a) $6050. (b) $6500. (c) $6550. (d) $7000.
(e) $7050 — 20. ________

21. A woman purchases 4 pounds of chopped meat at 80¢ per pound. What change does she receive from a ten dollar bill?
(a) $3.20 (b) $7.20 (c) $7.80 (d) $6.80
(e) $9.20 — 21. ________

22. One cup of condensed, sweetened milk weighs 11 ounces. How much milk will remain unused for a recipe requiring 2 cups of milk when a housewife opens 4 six ounce cans of condensed milk?
(a) 1 (b) 2 (c) 12 (d) 20 (e) 22 22. ________

23. 7 type p nails are worth 5 type q nails. 3 type p nails are worth x type q nails. What is the numerical value of x?
(a) $\frac{5}{7}$ (b) $1\frac{2}{3}$ (c) $2\frac{1}{7}$ (d) $3\frac{5}{7}$ (e) $4\frac{1}{5}$ 23. ________

24. A man runs y yards in m minutes. What is his rate in yards per hour?
(a) $\frac{y}{60m}$ (b) $\frac{m}{60y}$ (c) 60my (d) $\frac{60y}{m}$ (e) $\frac{60m}{y}$ 24. ________

25. Carol is running for class secretary. If there are 356 students who vote to choose the secretary, and a total of 5 candidates seeking the office, what is the least number of votes Carol must receive and still have more votes than any other candidate?
(a) 71 (b) 72 (c) 89 (d) 178 (e) 179 25. ________

26. A boy makes a diagram of a plane to a scale of $\frac{1}{400}$ of actual size. If the diagram of the plane is 9.6 inches in length, what is the length, in feet, of the plane?
(a) 96 (b) 320 (c) 384 (d) 960 (e) 3840 26. ________

27. $\dfrac{\frac{1}{3} - \frac{1}{12}}{\frac{1}{4}} = ?$
(a) $\frac{-4}{9}$ (b) 0 (c) $\frac{1}{4}$ (d) 1 (e) $\frac{5}{3}$ 27. ________

28. If the registration of a school increased from 300 to 1200, what is the percent of increase in registration?
(a) 30% (b) 90% (c) 75% (d) 300% (e) 400% 28. ________

29. The city of Cleveland has 154 Brownie troops taking part in a cookie sale. On a certain day, 40 troops have met their sales quota and 20 have sold no cookies at all. How many of the remaining troops must meet their quota so that approximately 65% of the troops have done so?
(a) 60 (b) 80 (c) 90 (d) 94 (e) 100 29. ________

30. If $\frac{x}{16}$ = .375, what is the numerical value of x?
(a) 2 (b) 3 (c) 4 (d) 5 (e) 6 30. ________

31. The batter which is 7 inches high in a loaf pan one and one quarter feet by 8 inches, is poured into a pan 13" by 20". What height will it reach in the larger pan?
(a) .27" (b) .31" (c) 1.7" (d) 3.2" (e) 4.6" 31. ________

32. Three times a number less seven is thirty-two. What is twice the number?
(a) 13 (b) 17 (c) 26 (d) 32 (e) 39 32. ________

33. In making drapes, Joan needs to cut 8 foot panels. How many panels may she cut from fabric that is y yards long?
(a) $\frac{y}{8}$ (b) $\frac{3y}{8}$ (c) 8y (d) $\frac{8y}{3}$ (e) 24y 33. ________

34. One gear rotates once every seven minutes, and another rotates once every five minutes. Both gears begin to rotate at the same time every (?) minutes?
(a) 6 (b) 12 (c) 17.5 (d) 35 (e) 70 34. ________

35. If the average ages of three people is 44 years, and if no one of them is less than 42 years old, what is the maximum age (in years) of any one person?
(a) 44 (b) 46 (c) 48 (d) 49 (e) 50 35. ________

36. If 2S seamstresses can sew 2D dresses in 2Wweeks, how many seamstresses would it take to sew 4D dresses in 4W weeks?
(a) S (b) 2S (c) 4S (d) 8S (e) 16S 36. ________

37. The secretary of a club can address the envelopes for a mailing in 40 minutes. Her younger sister who could do the entire job alone in 1 hour assists her. How long will it take to complete the job if both girls work ?
(a) 0.04 minutes (b) 0.4 minutes (c) 8 minutes (d) 24 minutes (e) 50 minutes 37. ________

38. An electric saw sells for $65.00. This price gives the hardware store owner a profit of 30% on his costs. What will be the new retail price if he cuts his profits to 10% of costs?
(a) $42. (b) $45.50 (c) $50.00 (d) $50.05
(e) $55.00

38. ________

39. Mrs. Stanley will be X years old 5 years hence. How old was she 5 years ago?
(a) X-5 (b) X+10 (c) X-10 (d) 5X-5
(e) 7X

39. ________

APPENDIX V

Post-hoc Tests

FEMALES

Treatments Means	Total Gp.Test	Female Gp.Test	Male Gp.Test
	14.36	15.61	12.93
	a	b	c

Ordered Differences

	c	a	b
c	-	1.43*	2.68*
a	-	-	1.25*
b	-	-	-

CV_{Tukey} $q_{.99}\ (3,\infty)\ \sqrt{MS_{res}/n}$ =

$4.12\ (\sqrt{1.55/148})$ =

$4.12\ (.102337)$ =

.4216

MALES

Treatments	Total Gp.Test	Female Gp.Test	Male Gp.Test
Means	17.90	13.98	16.02
	a	b	c

Ordered Differences

	b	c	a
b	-	2.04*	3.92*
c	-	-	1.88*
a	-	-	-

CV_{Tukey} $q_{.99}\ (3,\infty)\sqrt{MS_{res}/n}$ =

$4.12\ (\sqrt{4.25/1.85})$ =

$4.12\ (.151568)$ =

.6245

REFERENCES

Aiken, L.R. Personality correlates of attitude toward mathematics. Journal of Educational Research, 1963, 56, 476-480.

Aiken, L.R. Intellective variables and mathematics achievement: Directions for Research. Journal of School Psychology, 1971, 9, 201-212.

Aiken, L.R. Biodata correlates of attitudes toward mathematics in three age and two sex groups. School Science and Mathematics, 1972, 72, 386-395.

Aiken, L.R. and Dreger, R.M. The effect of attitudes on performance in mathematics. Journal of Educational Psychology, 1961, 52, 19-24.

Association of Womens Psychologists. Statement to the meeting of the American Psychological Association, Miami, 1970.

Atkinson, J.N. and Feather, N.T. A Theory of Achievement Motivation. New York: John Wiley & Sons, 1966.

Baumrind, D. From each according to her ability. School Review, 1972, 80, 2, 161-197.

Berry, P.C. An exploration of the inter-relationship among some non-intellective predictors of achievement in problem solving. Department of Psychology, Stanford University, 1958.

Brown,E.E. and Abel,T.L. Research in the teaching of elementary school mathematics. Arithmetic Teacher, 1965, 12, 547-549.

Carey,G.L. Sex differences in problem solving as a function of attitude differences. Journal of Abnormal and Social Psychology, 1958, 56, 256-260.

Carroll,J.B. Development of native language skills beyond the early years. The Learning of Language, Carroll E. Reed, Editor. New York: Appleton-Century-Crafts, 1971, 97-156.

Child,I.L., Potter,E.H., and Levine,E.L. Childrens' text-books and personality development: An explanation in the social psychology of education. Psychological Monographs, 60, 3, 1946, 1-54.

Cleary,T.A. Test bias: Prediction of grades of negro and white students in integrated colleges. Journal of Educational Measurement, 1968, 5, 115-124.

Coffman,W.E. Sex differences in responses to items in an aptitude test. 18th Yearbook, National Council on Measurement in Education, 1961, 117-124.

Coleman,R.J. An analysis of certain components of mathematical abilities and an attempt to predict mathematical achievement in a specific situation. Dissertation Abstracts, 1956, 16, 2062.

Darlington,R.B. Another look at "culture fairness." Journal of Educational Measurement, 1971, 8, 71-82.

Donlan,T.F. Content factors in sex differences on test questions. Paper presented at the meeting of the New England Research Organization, Boston, 1971.

Donlan,T.F. and Angoff,W.H. The scholastic aptitude test, in The College Board Admissions Testing Program: A Technical Report on Research and Development Activities Relating to the Scholastic Aptitude and Achievement Tests. W.H. Angoff (Ed.), College Entrance Examination Board, New York, 1971, 15-31.

Dwyer,C.A. Childrens' sex role standards and sex role identification and their relationship to achievement. Unpublished doctoral dissertation, University of California at Berkley, 1972.

Fennema,E. Mathematics learning and the sexes: A review. Paper presented at the meeting of the American Educational Research Association, New Orleans, 1973.

Fitzsimmons,S.J., Cheever,J., Leonard,E. and Mancunovich,D. School failures. Developmental Psychology, 1969, 1, 134-146.

Fremer,J. and Chandler,M. Sex differences on the scholastic aptitude test, in The College Board Admissions Testing Program: A Technical Report on Research and Development Activities Relating to the Scholastic Aptitude and Achievement Tests. W.H. Angoff (Ed.), College Entrance Examination Board, New York, 1971, 168-180.

Green,D.R. Racial and ethnic bias in test construction. CTB/McGraw Hill, 1972.

Gunderson,D.V. Sex roles in reading and literature. Paper presented at the meeting of the American Educational Research Association, Chicago, 1972.

Hall,E. The Silent Language. Greenwich, Connecticut: Fawcett Publishers, 1959.

Hartley,R. Sex role pressures and the socialization of the male child. Psychological Reports, 1959, 457, 268.

Hartley,R. and Klein,A. Sex role concepts among elementary school age girls. Paper presented at the meeting of the American Psychological Association, 1957.

Heimgartner,N.L. Selected mathematical abilities of beginning kindergarten children. Unpublished dissertation, Colorado State University. 1968.

Hervey,M.A. Childrens' responses to two types of multiplication problems. Arithmetic Teacher, 1966, 13, 268-292.

Horner,M.S. Sex differences in achievement motivation and performance in competitive and non-competitive situations. Unpublished doctoral dissertation, University, of Michigan, 1968.

Howe,F. Sexual stereotypes start early. Saturday Review. October 16, 1971, 76-82.

Husen,T.(Ed.) International Study of Achievement in Mathematics. A Comparison of Twelve Countries. New York: John Wiley & Sons, 1967.

Jarvis,O.T. Boy-girl ability differences in elementary school arithmetic. School Science and Mathematics, 1964, 64, 657-659.

Kagan,J. The emergence of sex differences. School Review, February 1972, 80, 2, 217-227.

Key,M.R. Linguistic behavior of male and female. Paper presented at the meeting of the American Dialect Society, New York, December 1970.

Leder,G.C. Sex differences in mathematics problem appeal as a function of problem context. Journal of Educational Research, 1974, 67, 351-353.

Lesser,F., Fifer,G., and Clark,D.H. Mental abilities of children from different social class and cultural groups. Monograph of the Society for Research in Child Development, 1965, 30, 1-115.

Lewis,M. Parents and children: Sex role development. School Review, February 1972, 80, 2, 229-239.

Lynn,D.B. Determinants of intellectual growth in women. School Review, February 1972, 80, 2, 241-260.

Maccoby,E.E. Differential socialization of boys and girls. Paper presented at the meeting of the American Psychological Association, Honolulu, 1972.

Maccoby,E.E. and Jacklin,C.N. Sex differences in intellectual functioning. Paper presented at the Educational Testing Service Invitational Conference on Testing Problems, New York, 1972.

Madden,R. New directions in the measurement of mathematical ability. Arithmetic Teacher, 1966, 13, 375-379.

Malcolm,A.H. Most common verb in school, study finds, is ...is. New York Times, September 4, 1971.

McMahon,I.D. Causal attributions and expectations of success: Age and sex differences. Unpublished dissertation, City University of New York, 1972.

Messick,S. What kind of difference does sex make? In Anderson,S.(Ed.), Sex Differences and Discrimination in Education. In print.

Milton,G.A. Five studies of the relation between sex role identification and achievement in problem solving. Technical Report #3, Department of Industrial Administration, Yale University, 1958.

Minuchin,P. The schooling of tomorrow's women. School Review, February 1972, 80, 2, 199-208.

Moss,H.A. Sex, age and state as determinants of mother-infant interaction. Merrill-Palmer Quarterly, 1973, 13, 19-36.

Newson,J., and Newson,E. Four years old in an urban community. Harmondworth, England: Pelican Books, 1968.

Nilsen,A.P. Women in children's literature. College English. 32, 918-926.

Preston,R.C. Reading achievement of German and American children. School and Society, 1962, 90, 350-354.

Quirk,T.J. and Medley,D.M. Race and subject matter influences on performance on general education items of the National Teacher Examinations. Paper presented at the meeting of the American Psychological Association, Honolulu, 1972.

Reese,H.W. Manifest anxiety and achievement test performance. Journal of Edcuational Psychology, 1961, 52, 132-135.

Sadker,M.P. Elements of sex bias in schools. Paper presented at the meeting of the American Educational Research Association, New Orleans, 1973.

Shine,A. Relationship between arithmetic achievement and item performance on the revised Stanford-Binet scale. Arithmetic Teacher, 1961, 8, 57-59.

Spilman,H. and Weiner,M. A study of student performance on arithmetic presented in verbal and non-verbal forms. Measurement and Evaluation in Guidance, 1975, 5, 446-449.

Sweeney,E.J. Sex differences in problem solving. Unpublished manuscript, Department of Psychology, Stanford University, 1953.

Terman,L.M. and Miles,C. Sex and Personality. New York: McGraw-Hill, 1936.

Thorndike,R.L. Concepts of culture fairness. Journal of Educational Measurement, 1971, 8, 63-70.

Tittle,C.K., McCarthy,K. and Steckler,J.F. Women and Educational Testing: A Selective Review of the Research Literature and Testing Practices. Princeton, New Jersey: Educational Testing Service, 1973.

Unkel,E. A study of the interaction of socioeconomic groups and sex factors with the discrepancy between anticipated achievement and actual achievement in elementary school mathematics. Arithmetic Teacher, 1966, 13, 66-670.

Very,P.S. Differential factor structures in mathematical ability. Genetic Psychology Monographs, 1967, 75, 169-207.

Watson,G. Psychological aspects of sex roles. Social Psychology: Issues and Insights, 1966, 427.

Wozencraft,M. Are boys better than girls in arithmetic? Arithmetic Teacher, 1963, 10, 486-490.

Wrigley, . The factorial nature of ability in elementary mathematics. British Journal of Educational Psychology, 1958, 28, 61-78.

Yarrow,L.J., Rubenstein,J.L. and Pedersen,F.A. Divensions of early stimulation: differential effects on infant development. Paper presented at the meeting of the Society for Research in Child Development, 1971.