DOCUMENT RESUME

ED 269 244 SE 046 560

AUTHOR Marshall, Sandra P.
TITLE Errors in Processing Mathematical Information: A Cross-Sectional and Longitudinal Study of Individual and Sex Differences. Final Report.
SPONS AGENCY National Inst. of Education (ED), Washington, DC.
PUB DATE Apr 86
GRANT NIE-G-83-0048
NOTE 50p.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Cognitive Processes; Cross Sectional Studies; Educational Research; Elementary Education; *Elementary School Mathematics; *Error Patterns; Grade 3 Grade 6; *Longitudinal Studies; *Mathematics Achievement; *Sex Differences
IDENTIFIERS California; *Mathematics Education Research

ABSTRACT

This study investigated the rates of successful mathematical performance and errors of information processing in third-grade children in California and continued an investigation of these factors in sixth-grade children. The objectives were to: (1) identify areas of mathematics in which the children had recognizable strengths and weaknesses, (2) classify characteristic areas according to information-processing theory, and (3) relate the errors made by third graders in 1980 with errors made by the same children as sixth graders in 1983. Approximately 25,000 children in each grade were tested for each year using data from the Survey of Basic Skills in the California Assessment Program. The tests and population are first briefly discussed, followed by the results of the analyses. For each grade level, discussion focuses on: correct performance by categories --computation, counting/number property, word problems (using operations), visual problems, geometry/measurement problems, and nontraditional story problems; the most difficult items; scores on matched items (computation and applications); and errors. Then the longitudinal analysis is dicussed. Girls appear stronger than boys in grade 3 and boys were stronger than girls in grade 6. References and sample items are included in the appendix. (MNS)

Errors in Processing Mathematical Information: A Cross-Sectional and Longitudinal Study of Individual and Sex Differences

NIE-G-83-0048
Final Report

Sandra P. Marshall
Department of Psychology
San Diego State University
San Diego, CA 92182

April 1986

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

The present project was an investigation of rates of successful performance and errors of information-processing in third-grade children and a continuation of investigation of these issues in sixth-grade children. Three objectives were specified for the project. Each is described below, together with a brief statement describing success in reaching the objective.

Objectives

The objectives were:

(1) To identify areas of mathematics in which third-grade and sixth-grade girls and boys have recognizable strengths and weaknesses.

At the third grade, girls performed significantly better than boys in the areas of arithmetic computations, principles of counting, and nonstandard problem solving. At the sixth grade, boys had significantly more success than girls in solving geometry/measurement problems and traditional word problems. Girls maintained their advantage for arithmetic computations.

(2) To classify characteristic errors made by either sex at the two grades according to information-processing theory.

At both grades, boys were more likely than girls to make errors related to usage of erroneous arithmetic rules, including errors of number fact, errors in using algorithms, and errors of confusion with horizontal problems. They were also more likely to select the opposite semantic category (e.g., respond with the greatest rather than the least value when the least was required). Girls were more likely than boys to make errors of association, e.g., focusing on particular words in the problems and using inappropriate rules such as adding all numbers in the problem. Both boys and girls made errors related to attention, with boys more likely to make careless errors of transcription and girls more likely to omit a step of the solution.

(3) To relate the errors made by a large sample of third-grade children in 1980 with the errors made by the same children as sixth graders in 1983.

There is substantial improvement in children's mathematical performance from third to sixth grade. However, a large number of children who failed to solve particular problems at the third grade remained unable to solve similar problems at the sixth grade. Girls were more likely than boys to be incorrect on items at both grades.

The data studied were responses to standardized achievement tests taken by all California third and sixth grade children enrolled in public schools. Approximately 250,000 children at each grade were tested for each year studied. The data were gathered by the California Assessment Program of the California Department of Education.

This research provides new information about the nature of errors made by elementary school children. Children's responses were examined in the context of cognitive skills and information processing. A more usual method of research has been to study only correct performance within narrowly defined subfields of mathematics such as geometry or arithmetic. Emphasis in this study was on cognitive behaviors --correct and incorrect-- that apply over many different subfields. An advantage of a large study such as the one carried out here is that many subfields of third and sixth grade mathematics could be studied simultaneously. The evaluation of a large number of children's responses to a large number of items provides information about similarities and differences in children's problem solving at two important ages.

TEST INSTRUMENTS AND POPULATION

Responses of third-grade and sixth-grade children to grade-level standardized tests were examined. The tests are the Surveys of Basic Skills, Grades 3 and 6, administered annually to all third-grade and sixth-grade children enrolled in public schools in California. The tests were developed by and are administered under the California Assessment Program (CAP) of the California Department of Education. The tests assess reading, written expression, and mathematics performance. Additional details may be found in the California Assessment Program Annual Report (1983).

These tests were designed to assess the average performance of children at school, school district, and state levels. Individual results are not released to the schools or to the students. A variety of items are included in the tests, and the objective is to evaluate a large number of separate concepts identified from the curricula of third-grade and sixth-grade mathematics.

The Third-Grade Test

The Survey of Basic Skills, Grade 3, contains 360 mathematics items. There are 30 distinct test forms, and each contains 12 math items. Each student responds to a single form. The tests are not equally difficult, and the items on each form usually test different concepts. Seven areas of mathematics are evaluated by the Survey:

arithmetic operations	155 items
counting and place value	45 items
number properties	45 items
measurement	40 items
geometry	30 items
patterns and graphs	30 items
nontraditional word problems	15 items

In all but the last category, at least two types of problems, "skills" and "applications", test the concepts. Skill items are simple computations. Applications are word problems requiring the identification and use of skills for solution.

An additional feature of the Survey is the inclusion of matched pairs of skill and application items using the same numerical values and having the same set of distracters. For example, the items below are matched:

$$\begin{array}{r} 78 \\ +45 \\ \hline \end{array}$$

( ) 33 ( ) 123
( ) 133 ( ) 1113

Jenny baked 45 cookies.
Then she baked 78 more.
How many cookies did she bake?

( ) 33	( ) 123
( ) 133	( ) 1113

There are 32 pairs of matched items on the test.

The mathematics section of the third-grade test was first administered in May 1979, and has been given every spring thereafter. In this research project, we evaluated responses from the 1980 administration.

The Sixth-Grade Test

The _Survey of Basic Skills, Grade 6_, is similar in design to the third-grade test. The first test of the sixth-grade level was administered from 1975 through 1981 and contained 160 mathematics items in essentially the same content areas as those described for the third-grade test. It also contained items of probability and statistics. The test was revised and expanded in 1981. It currently contains 480 items distributed in the following categories:

arithmetic operations	145 items
counting and place value	4[illegible] items
number properties	50 items
measurement	58 items
geometry	40 items
equations and coordinate graphs	42 items
tables and charts	30 items
probability and statistics	23 items
nontraditional word problems	52 items

There are 12 pairs of matched items on the sixth-grade test. For example:

0.5 + 0.03 =

( ) 0.008
( ) 0.08
( ) 0.53
( ) 0.8

A paper clip weighs 0.5 grams. A piece of paper weighs 0.03 gram. How much would the paper and the paper clip weigh?

( ) 0.008
( ) 0.08
( ) 0.53
( ) 0.8

The revised test was first administered in May 1982 and has been given annually since that time. Individual student responses to the 1983 administration were used in this project.

It should be noted that the children responding as third graders in 1980 were sixth graders in 1983. Therefore, the responses to the sixth-grade test in May 1983 are doubly valuable: they provide information about sixth-grade problem solving in general and they also contain longitudinal information about the development of problem-solving skills and use of cognitive processes from the third to the sixth grade.

Population

Every third-grade and sixth-grade child enrolled in public school in California responds to the standardized tests described above. Approximately 250,000 - 300,000 children at each grade are tested annually. The population varies by sex, by age in months, by the primary language spoken at home, by geographic location, and by socioeconomic status. These student characteristics are collected for each individual together with item responses.

Responses from all students were examined in the initial comparisons. The results of these investigations are reported in the second and third sections of this report. A subpopulation was identified for the longitudinal study, discussed in section four. For this subset of data, attention was restricted to children enrolled in the same school at grades three and six. This enrollment information is routinely gathered by the California Assessment Program when the sixth-grade tests are administered.

The California Assessment Program makes student identification through the personal characteristics described above, namely, sex, birthdate, primary language, and ethnicity. CAP does not record student names (since individual test scores are not released). Therefore, the process of matching third-grade and sixth-grade responses for individuals was based upon these same personal characteristics. For each school, the third- and sixth-grade individuals were matched according to sex and birthdate. The estimates of primary language and socioeconomic status were not used as matching variables. These responses were estimates made by the teachers at each grade. It was feared that teachers from grade to grade might differ in their estimation of children's socioeconomic background and of the language spoken most often at home. It is also possible that one or both of these variables might have changed within the three year period from third to sixth grade. The variables of school, birthdate, and sex were invariant over this period.

We were able to locate full test data at both third and sixth grade for roughly 100,000 students. Our final subset of data contains responses from children enrolled in elementary schools that span third through sixth grade (at least). In the initial population of 300,000 third grade students, about 150,000 students were in elementary schools that covered only kindergarten through fourth or fifth grades. These students then moved to middle schools containing grades six through eight. We had no means of matching feeder elementary schools with middle schools and thus were unable to follow these children. The remaining students not in our matched subset were students who failed to give full demographic information at one or both of the test administrations or students in the same school having identical personal characteristics. This final criterion meant that identical twins or fraternal twins of the same sex were excluded, since they manifested identical demographic data.

RESULTS OF THE THIRD-GRADE ANALYSES

For the analyses described in this and subsequent sections of the report, the test items from the Survey of Basic Skills: Grades 3 and 6 were evaluated according to six categories that were common to both tests. Consequently, some of the items (e.g., probability items from the sixth-grade test) were not analyzed because they occurred only at a single grade level. The categories used here are given in Table .

Table 1

Description of Items

CATEGORY LABELS	FREQUENCY OF OCCURRENCE: THIRD GRADE
(1) Computations	115
(2) Counting and Properties of Numbers	87
(3) Word Problems	44
(4) Visual Problems	57
(5) Geometry and Measurement Problems	39
(6) Nontraditional Story Problems	18

Some of these categories differ from those used by the California Assessment Program. The category of computations refers to problems given in traditional equation or expression form for which the student must carry out the indicated operation(s). Counting and number properties items are those that require the student to demonstrate knowledge of concepts such as even/odd, series, and place value. Word problems are traditional story problems in which one or more arithmetic operations are embedded. The category of visual problems contains all problems with a visual component, such as charts, graphs, or diagrams (excluding problems of identification of geometrical shapes). Geometry and measurement problems are discussed as a single category because of the overlap between these two types of problems in elementary school. The final category contains nontraditional story problems that require the student to make a noncomputational response. For example, problems in this category may request identification of the facts required to solve the problem, identification of a restatement of the problem, or recognition of a similar problem. Examples from each category may be found in Appendix A.

Correct Performance

Categories of Items

In general, the third graders performed quite well on the first administration of the Survey of Basic Skills: Grade 3. A summary of their overall rate of success by sex is given in Table 2.

Table 2

Correct Performance on the SURVEY OF BASIC SKILLS: GRADE 3

Area	Percent Correct Boys	Girls
Computation	72.12	74.20
Counting/Number Property	71.45	72.88
Word	61.94	64.65
Visual	75.48	75.93
Geometry/Measurement	65.86	66.45
Nontraditional	66.63	69.50

For both boys and girls, word problems were the most difficult items of the test and visual problems were the easiest items. A rank order of the categories is identical for the sexes. From easiest to most difficult they are: visual, computation, counting and number property, nontraditional, geometry and measurement, and word problems.

Comparisons were made to determine whether the probabilities of success for each category differed within each sex. That is, were boys equally likely to succeed on computation or word problems, or were there statistically significant differences between the rates of .7229 and .6494? For boys, the rates of success over all categories differed significantly from each other with three exceptions. Counting and computation items showed no difference, and the two categories of word problems and nontraditional problems did not differ from geometry/measurement items. There were significant differences between word problems and nontraditional items. For girls, all categories were significantly different from each other.

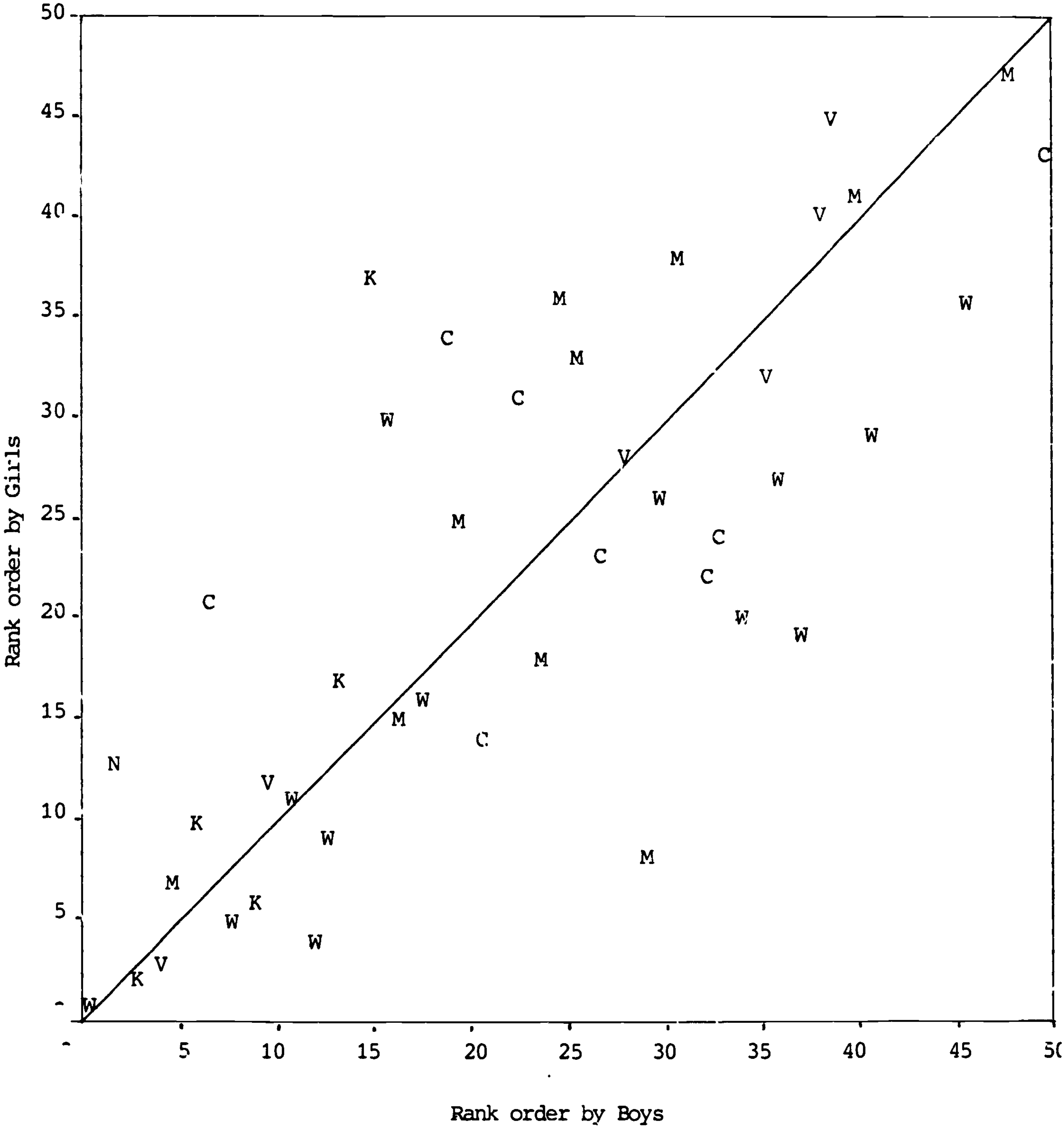
There appear to be two patterns for boys' and girls' success rates over these categories. For boys, there are two groups of items, one group containing word problems, nontraditional problems, and geometry/measurement significantly easier for boys than the other three categories. A different pattern emerges for girls. Like boys, they found computations, counting/number properties, and visual items to be easiest. However, the category of nontraditional items does not group with the other two categories of word problems and geometry/measurement. Instead, these latter two categories form a difficult group similar to that observed for boys. The nontraditional items comprise a third group of intermediate difficulty.

Although they were identical in the rank order of category difficulty, boys and girls differed in the degree of difficulty associated with each category. For each category, the probability of success by boys was compared with that of girls. Three of these comparisons were statistically significant beyond the usual .05 probability level: girls had higher levels of success on computations, items of counting and number properties, and nontraditional problems. They were also marginally better in solving items of measurement/geometry and visual items. Boys were not significantly more successful than girls over any category, although they demonstrated slightly higher rates of success for the word problems.

<u>The Most Difficult Items</u>

A second analysis provides information about which particular items were most difficult for boys and girls. The 50 items having the lowest p-values were identified for each sex (i.e., the most difficult items). As one might expect, a large majority of those that caused difficulty for one sex also caused similar trouble for the other sex. However, there were seven items that appeared on the most difficult list for boys that did not have similar difficulty for girls. Thus, on 14 percent of the most difficult items, boys and girls did not agree. Six of these seven items were arithmetic computations; the seventh was a nontraditional item requiring identification of the question asked in the problem. Three of the six computational items were multidigit subtraction items, one was a simple multiplication problem, and the remaining two were horizontal multiplication problems involving only single digits.

The items that were difficult for girls but not for boys were three word problems, one visual problem, and two items requiring multiplication of 10 or 100. This last weakness has been noted before (CAP, 1981). While girls are consistently better able than boys to answer problems of simple arithmetic computation, they have difficulty when the numbers are multiples of 10. We have no explanation for this finding.



C Computation
K Counting/Number Property
W Word
V Visual
M Geometry/Measurement
N Nontraditional

Figure 1: A Comparison of the Most Difficult Items for Boys and Girls: Grade 3.

The results of comparing the most difficult items for each sex are consistent with the overall comparisons of categories. Boys have more difficulty than girls with computational items and girls have more difficulty than boys with word problems.

There remained 43 common items of difficulty for boys and girls. Each of these had an assigned rank from the above listing of the 50 most difficult items for boys and girls. These ranks were compared using the standard Pearson product-moment correlation coefficient. The degree of similarity for these two sets of ranks may be seen in the correlation value of .77. This value indicates that in general boys and girls found the same items to have similar relative difficulty. However, the value of .77 also indicates that there are a substantial number of other items in the common set which have different ranks.

Of the 43 common items falling into the 50 most difficult items for both boys and girls, 12 differed in rank order by more than 10 places. For example, the item that was the second most difficult item for boys had a rank of 13 for girls. One that held a rank of 15 for boys (with 1 being the most difficult) ranked only 37 with girls. The eighth most difficult item for girls ranked 29 with boys. Thus, the high correlation masks some wide differences in relative difficulty. Again, the item type reflects the findings reported above. Four of the six items which girls found to be substantially more difficult were word problems. Two of the six considered to be difficult by boys were computational items.

Figure 1 contains a summary of these findings. The 43 items of common difficulty are plotted in this figure. The rank for boys is on the ordinate and that for girls is on the abscissa. Each item is identified in the plot by its category of Table 1. Items falling in the northwest quadrant of the figure were relatively more difficult for girls and less difficult for boys. Items in the southeast quadrant have the reverse pattern; these are more difficult for boys and less difficult for girls. It is evident that most of the items in the former are word problems. There is no clear pattern for boys; items from each category are in this quadrant.

Matched Items

There are 25 pairs of matched items on the this test. Each pair contains a computation (or skill) item and an application item requiring the same skill. For this set of items, we get the results of Table 3.

Table 3

Average Correct Performance of Boys and Girls on Matched Items: Third Grade

	Computations	Applications
Boys	80.76	73.49
Girls	80.91	72.67

For both sexes, the difference in performance on computations and applications is statistically significant ($p < .05$). There are no differences between boys and girls on either type of item.

The relationship between computation and application performance is more clearly observed in Figure 2. Each pair of items is plotted. Boys' performance is indicated by the symbol X and girls' performance is given by O. As can be seen in the figure, there appears to be a linear relationship for boys and for girls between performance on the two types of items. Regression equations for each group were developed. For boys, the equation for predicting application performance (A) from computation performance (C) is:

$$A = 1.116C - 16.656.$$

For girls, the corresponding equation is:

$$A = 1.096C - 16.000.$$

Tests of both regression equations were significant, indicating that the regression of applications on computations is significantly different from zero ($F = 43.00$, $df = 1,23$; $F = 28.60$, $df = 1,23$; $p < .001$ for both). A comparison of the regression coefficients in the two equations was nonsignificant. The relationship between computations and applications is the same for boys and for girls.

These statistical tests suggest that for all children, there is a reasonably constant relationship between performance on computation items and performance on corresponding application items. The difference between the two is large. Performance on both is measured on the same scale, percentage correct. As can be seen in the comparison of means in Table 3, performance on applications lags behind performance on computations by approximately eight percentage points. One concludes that students know how to compute successfully many different types of problems (computations), but do not know when these computations are appropriate (applications).

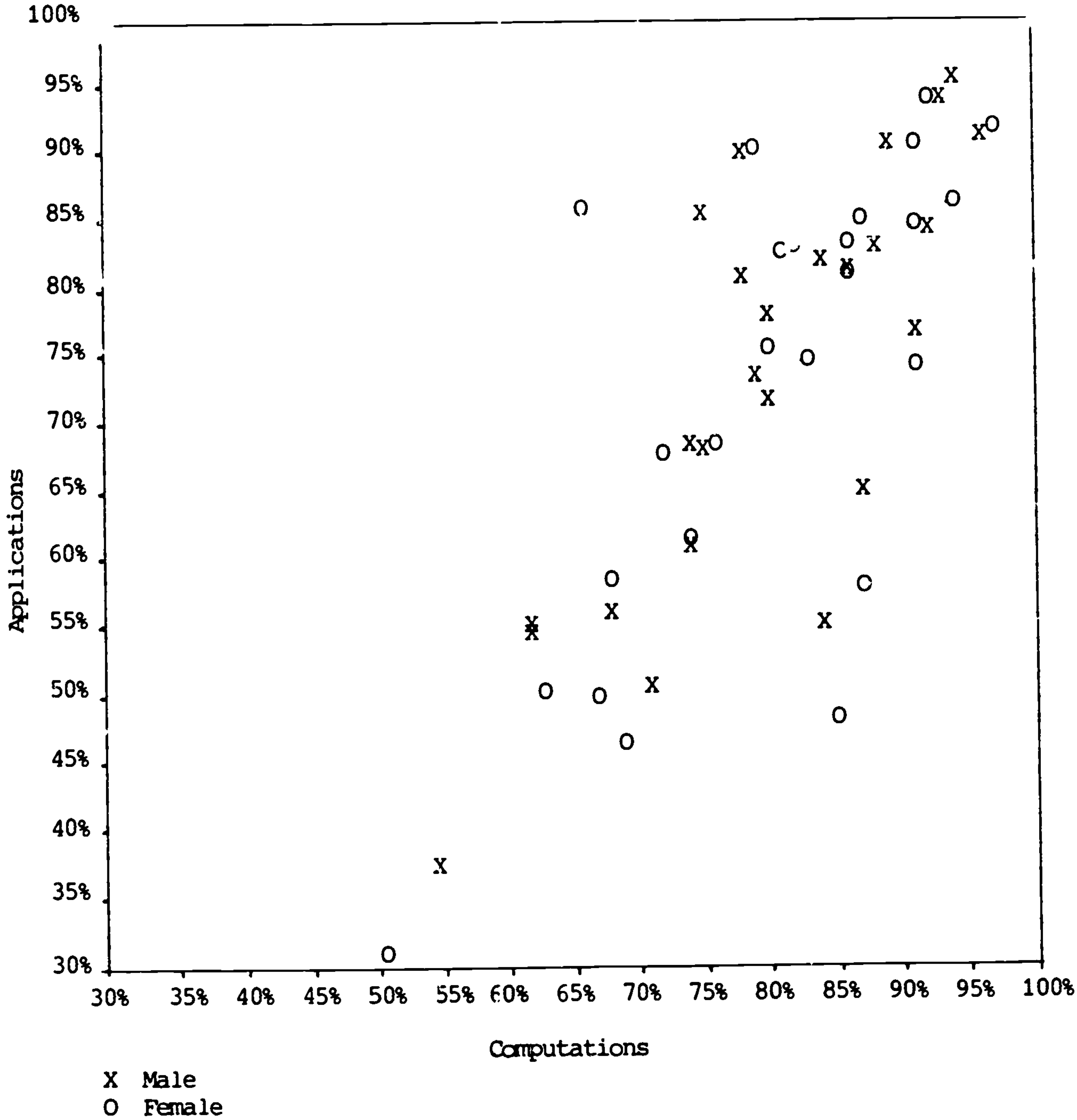


Figure 2: Performance of Boys and Girls on Matched Items of Computations and Applications: Grade3.

Analyses of Errors

In previous research, we have found that boys and girls have tendencies to make different types of errors on mathematics problems. In an earlier project funded by the National Institute of Education (Grant No. NIE-G-80-0095), I developed a classification of errors based upon the cognitive processes used (Marshall, 1982; 1983). The classification had the following categories:

I. Language
II. Spatial Understanding
III. Mastery
IV. Association
V. Irrelevant Rules
VI. Erroneous Rules

This classification was used here as well, but the category of "Irrelevant Rules" was replaced by a category of "Lack of Attention."

Under each of the six general types of errors listed above fall many distinct hypotheses about children's performance. In addition to the hypotheses formulated and tested in the earlier project, many new hypotheses have been proposed and evaluated here. In the original classification, many hypotheses about errors could not be tested because of the limited number of test items (160) and the inappropriate set of distracters for many items. In the current research, there are 360 items at the third grade and 480 items at the sixth grade. Most of these items have reasonable and usable distracters.

The types of distracters evaluated in separate hypotheses are given in Table 4 together with their parent category. The list of errors in Table 4 is a result of theoretical considerations and empirical assessibility. We began with an assessment of the categories and types of errors studied in the previous NIE project (Marshall, 1982). We then examined all items on the third-grade and sixth-grade tests. Our examination yielded several additional errors that could be evaluated, particularly in the category of erroneous rules. There are undoubtedly other erroneous rules that students use in solving mathematics problems. The ones evaluated here are those with distracters corresponding to the errors.

Table 4

Distracter Analysis

TYPES OF DISTRACTERS	NO. OF ITEMS HAVING THIS DISTRACTER	NO. OF TIMES BOYS MAKE ERROR MORE THAN GIRLS	NO. OF TIMES GIRLS MAKE ERROR MORE THAN BOYS	
I. LANGUAGE ERRORS	22	12	10	
A. Literal Translations	6	1	5	*
B. Opposites	16	11	5	*
II. SPATIAL UNDERSTANDING	14	10	4	**
A. Spatial Reversals	14	10	4	**
III. MASTERY	118	59	59	
A. Wrong Operation	118	59	59	
IV. ASSOCIATION	25	3	22	***
A. Key Words	11	1	10	***
B. Number Patterns	14	2	12	***
V. ERRONEOUS RULES	132	106	26	***
A. Subtract Small from Large	16	13	3	***
B. Right to Left Reversals	10	8	2	**
C. Add All Digits	18	14	4	***
D. Expand Columns	14	13	1	***
E. Mix Two Operations	14	5	9	
F. Borrowing Errors	15	12	3	***
G. Carrying Errors	17	17	0	***
H. Concatenations	28	24	4	***
VI. LACK OF ATTENTION	76	34	42	
A. Omit a Step	7	4	3	
B. Careless Transcription	21	14	7	**
C. Lack of Perseverance	15	3	12	***
D. Interference: Series	16	6	10	
E. Partial Reading	17	7	10	

* $.10 < p < .20$
** $.05 < p < .10$
*** $p < .05$

Description of the Different Errors and Their Corresponding Distracters.

Language Errors. There are two errors under the category of "Language", (a) literal translations and (b) opposites. Literal translation refers to the errors made in translating words directly into numbers (and vice versa) without regard to place value information. For example, an error of this type would be the response of 30046 to the question: "How would you write three hundred forty-six".

Errors of opposites refer to confusions in semantic understanding. An example is the response of the least value when the largest is requested.

Errors of Spatial Understanding. Only a single error was evaluated at the third grade, the error of spatial reversals. Spatial reversals are responses that confuse spatial orientation of top and bottom, left and right.

Errors of Mastery. There is a single type of error in this category, application of an incorrect arithmetic operation.

Errors of Erroneous Rules. At the third grade, this is the largest category of errors. Apparently many children have not yet mastered the correct procedures for carrying out arithmetic operations. The first error of Table 4 is that of subtracting the smallest value from the largest value without regard for where the values are placed in the problem. Thus, 24 - 18 = 14, by this erroneous rule.

The second error of this class is the reversal of left to right in placing numbers. For example, when asked to write the number three hundred forty-two, a student might respond with 243.

The third error is that of adding all digits in the problem. Given the addition problem of 25 + 16, a student following this rule sums the digits 2, 5, 1, and 6 for a response of 14.

The error labeled "expanding columns" refers to addition of two or more columns as if each column were independent. Under this rule, one gets the following response to the addition problem of:

```
  76
  89
 ---
1515
```

The error of mixing two operations occurs when a student begins to apply one operation such as addition and then switches to a second procedure such as multiplication within the same problem.

The next two errors are those of borrowing and carrying. Students either omit these procedures entirely or apply them to inappropriate columns.

The final third-grade error of erroneous rules is that of concatenation. This is an error in which the student simply concatenates the digits present in the problem. Thus, for the addition of 24 + 56, a student using this rule responds 2456.

Errors from Lack of Attention. Errors in this category are those that appear to result from lapses in attention or from attention to inappropriate elements of a problem. The first of these is the restricted attention to key words in the problem. At the third-grade level, this error is the association of the key words "How many" with the operation of addition. This pairing has been previously verbalized by sixth-grade students in an interview setting (Marshall, 1982).

The second error of the category results presumably from lapses in attention. This is the error of careless transcription of numbers, such as responding 163 rather than 136.

The error of lack of perseverance refers to failure to apply the same rules or procedures repeatedly as needed in a problem. Typically, an error of this type reflects failure to carry out the final step in a problem when that step is identical to the one just performed.

Another error of attention is the interference of a known pattern with the one being used in a particular problem. Most of the problems at the third grade are problems in identifying series of numbers. Given a series such as: 2,4,6,__?, an error of interference would be the response of 7, in which typical counting behavior interfers with the process of counting by 2.

The final error of this category is labeled partial reading. Errors of this type correspond to attention only to part of the stated problem. For example, given the problem:

> There were 3 dogs. Each dog had 3 puppies.
> How many puppies were there?

An error of partial reading would be 3 puppies.

Table 4 contains the results of statistical tests comparing the performance of boys and girls on each type of error. Consider first the individual types of errors. There are 19 in total. For each type of error, we recorded the number of times the error could be made in the 360 items, the number of times boys were more likely than girls to make the error, and the corresponding number of times that girls were more likely to err. The null hypothesis for each error was that boys and girls were equally likely to err. Of the 19 distinct tests recorded in Table 4, nine yielded probabilities smaller than the usual .05 level of significance. These are indicated by *** in the table. Thus, boys and girls were NOT equally likely to make the same errors on roughly half of the errors identified. The results of five additional tests were marginally significant (between .05 and .10).

Table 4 also gives the aggregation of errors within parent categories. Comparisons of these are revealing. In particular, girls were clearly more likely than boys to make errors of association while boys were much more likely than girls to make errors using erroneous rules. Further, if one excludes the error of careless transcription from the category of attention, girls were also significantly more likely than boys to make errors related to attention.

RESULTS OF THE SIXTH-GRADE ANALYSES

Most of the sixth-grade items were classified by the same categories as those of the third grade. Two additional categories are needed at the sixth grade, probability and algebra. The distribution of items over categories is given in Table 5.

Table 5

Description of Items

CATEGORY	FREQUENCY OF OCCURRENCE: SIXTH GRADE
(1) Computations	86
(2) Counting and Properties of Numbers	104
(3) Word Problems	63
(4) Visual Problems	73
(5) Geometry and Measurement Problems	77
(6) Nontraditional Problems	53
(7) Probability	16
(8) Algebra	8

The last two categories will not be discussed further since they have no counterpart at the third grade. Items from these categories were excluded from all analyses.

Correct Performance

The level of difficulty of the sixth-grade test differs from that of the third-grade test. The mean percentages at the third grade ranged from 64.94% to 75.93%. The range at the sixth grade is approximately the same size but slightly lower, from 57.48% to 68.17%. Table 6 contains a summary of the correct performance by boys and girls.

As in the third grade, the most difficult items here are word problems. Both boys and girls found them to be significantly more difficult than other items. However, girls and boys differed in the rank order of category difficulty. The order of difficulty for boys from easiest to hardest is: computation, visual, counting/number properties, geometry/measurement, nontraditional and word problems. For girls, the order is: computation, counting/number properties, visual, nontraditional, geometry/measurement, and word problems. With the exception of visual problems, girls maintained the order found at the third grade. Although the rank order for boys indicates a shift in the difficulty of visual and counting/number property items and a similar shift between geometry/measurement and nontraditional items, inspection of Table 6 shows that these percentages are very close. There is also little difference for girls between the categories of visual and counting/number properties. However, there is substantial difference between nontraditional and geometry/measurement problems, with the former being less difficult.

Table 6

Correct Performance on the SURVEY OF BASIC SKILLS: GRADE 6

Area	Percent Correct Boys	Girls
Computation	65.84	68.17
Counting/Number Property	65.66	65.47
Word	59.19	57.48
Visual	65.66	65.32
Geometry/Measurement	63.90	61.89
Nontraditional	63.74	64.47

Comparisons among categories for each sex yield different patterns of difficulty for boys and for girls. The categories of visual, counting/number properties, and computations have essentially the same difficulty for boys, and the categories of nontraditional and geometry/measurement items are also indistinguishable. These last two are significantly more difficult that the other three. Finally, word problems are significantly more difficult than the pair of nontraditional and geometry/measurement items. The significance level used for all tests of categories was .05.

Girls demonstrated a different pattern of item difficulty. Computation items were significantly easier than any other category. The three categories of nontraditional, counting/number properties, and visual items had essentially the same level of difficulty. These were significantly easier than the category of geometry/measurement, and the latter was itself significantly less difficult than the most difficult category of word problems. Again, the significance level was .05.

Comparisons were also made between boys and girls for each category. Recall that at the third grade, girls scored significantly better than boys on computations, counting/number properties, and nontraditional items. At the sixth grade, girls continued to outperform boys on items of computation but lost ground in four areas. First, they lost the advantage demonstrated at third grade in the areas of counting/number properties and nontraditional items.

There was no difference in correct performance of boys and girls at the sixth grade. Second, boys moved from approximately equal performance at third grade with girls in the areas of word problems and geometry/measurement to superior performance in these two areas at the sixth grade. These differences were statistically significant ($p < .05$).

The Most Difficult Items

An analysis similar to that described in the previous section was carried out with the sixth-grade data. The 50 items having the lowest probability of being answered correctly were identified for boys and for girls. As was found at the earlier grade, there is a large overlap in the items that are difficult for students of both genders. At the third grade, there were 43 common items. At the sixth grade, there are 40 common items. The rank correlation for these 40 items was .79, essentially the same value as before.

Of interest are the items that were difficult for one gender and less difficult for the other. Of the 10 items included in the difficulty list for boys but not for girls are five computations, two geometry/measurement items, two visual items, and one word problem. This is consistent with the findings at the third grade: Boys have more difficulty with computations than do girls. The corresponding list of 10 items that are more difficult for girls than for boys contains two items each of geometry/measurement, word problems, counting/properties of numbers, and visual items, and a single computation.

An examination of the differences between the rankings for boys' performance and girls' performance on the 40 common items revealed that for 2 items, the rank difference between boys and girls exceeded 15. For 9 items, the rank difference fell between 10 and 15 points; for an additional 9 items, the difference was between 5 and 10. The remaining 20 items were ranked essentially the same by boys and girls (i.e., were within 5 ranks).

It is useful to examine those items with rank difference greater than 5. There are 20 such items. Seven of them (35%) are geometry/measurement items. Girls had lower ranks on six of the seven (lower rank indicates greater difficulty). Four of the items were counting/properties of numbers. Boys found all of these to be more difficult than did girls. Boys' ranks for the 2 word problems and 2 nontraditional problems were also lower than the ranks from girls' performance. There were no visual items with rank difference greater than 5. The computation items showed no pattern.

Figure 3 shows the 40 items that were most difficult for both boys and girls. As before, those items in the northwest quadrant were relatively more difficult for girls than for boys; those in the southeast were relatively more difficult for boys. Most of the items in the former are geometry/measurement problems; many of those in the latter are counting or computation items. Again, these observations are consistent with previous findings related to girls' and boys' performance.

Matched Items

There are 11 pairs of matched items. As on the third-grade test, these are matched computations and applications using the same skills and the same numerical values.

Table 7

Average Correct Performance of Boys and Girls on Matched Items: Sixth Grade

	Computations	Applications
Boys	70.51	63.48
Girls	72.65	63.63

At the sixth grade, girls perform significantly better than boys on computations, but there is no difference in performance on applications. This means, in effect, that there is a larger discrepancy between girls' performance on the two types of items than between boys' performance on the two. This finding is supported by previous research that found this discrepancy to be statistically significant (Marshall, 1984).

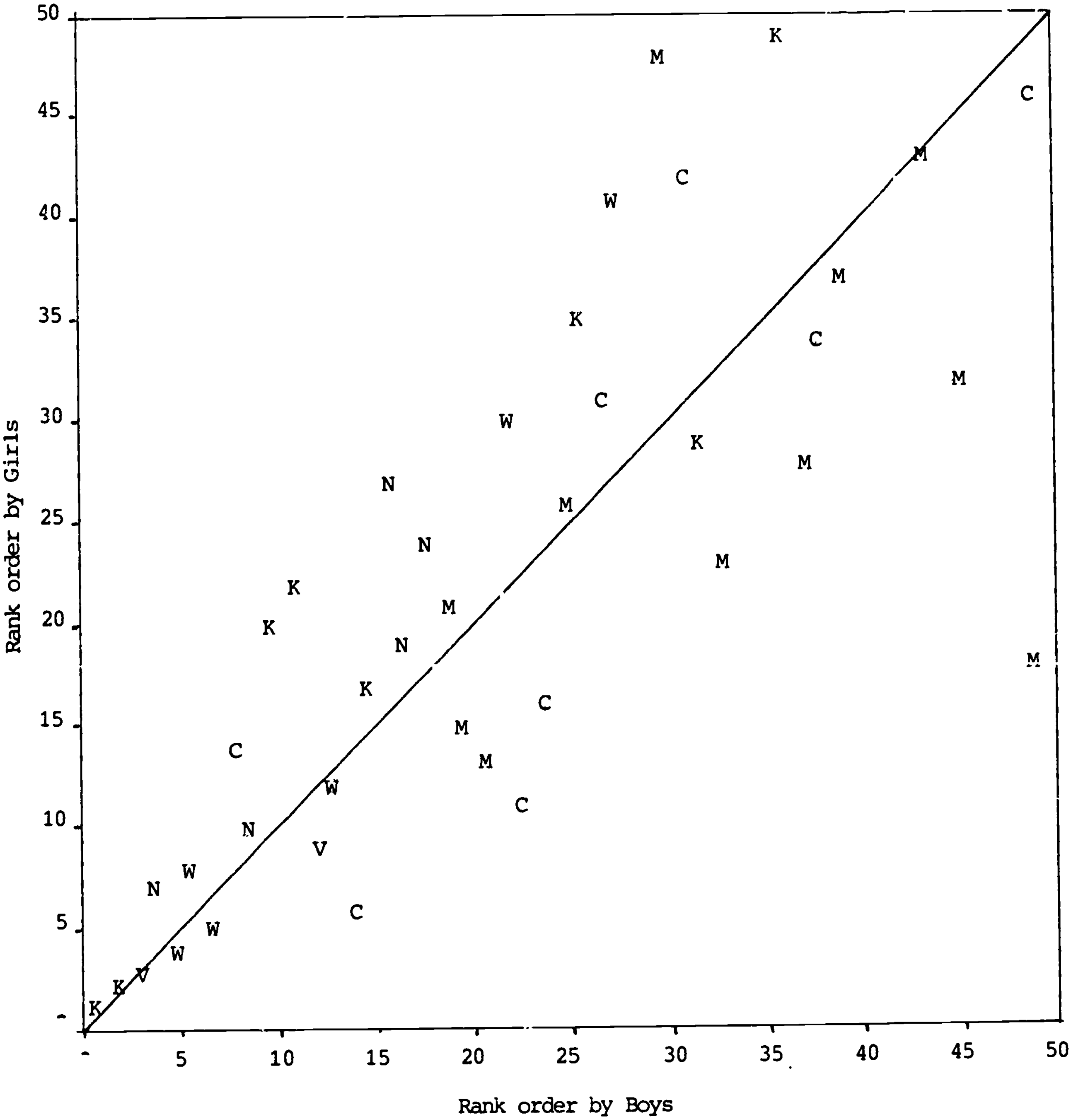
The matched-item data is plotted in Figure 4. There are fewer matched items at sixth grade than at third, but the linear trend is nonetheless apparent. The regression equations for predicting application performance from computation performance for boys and girls are:

$$A = 0.638C + 18.482$$

and

$$A = 0.664C + 15.414$$

respectively. The difference in intercepts corresponds to the difference in mean performance discussed above. As in the third grade results, tests of the regression equations are significant ($F = 21.58$, $df = 1,9$; $f = 20.65$, $df = 1,9$; $p < .001$), and the two coefficients are not significantly different from each other.



C Computation
K Counting/Number Property
W Word
V Visual
M Geometry/Measurement
N Nontraditional

Figure 3: A Comparison of the Most Difficult Items for Boys and Girls: Grade 6.

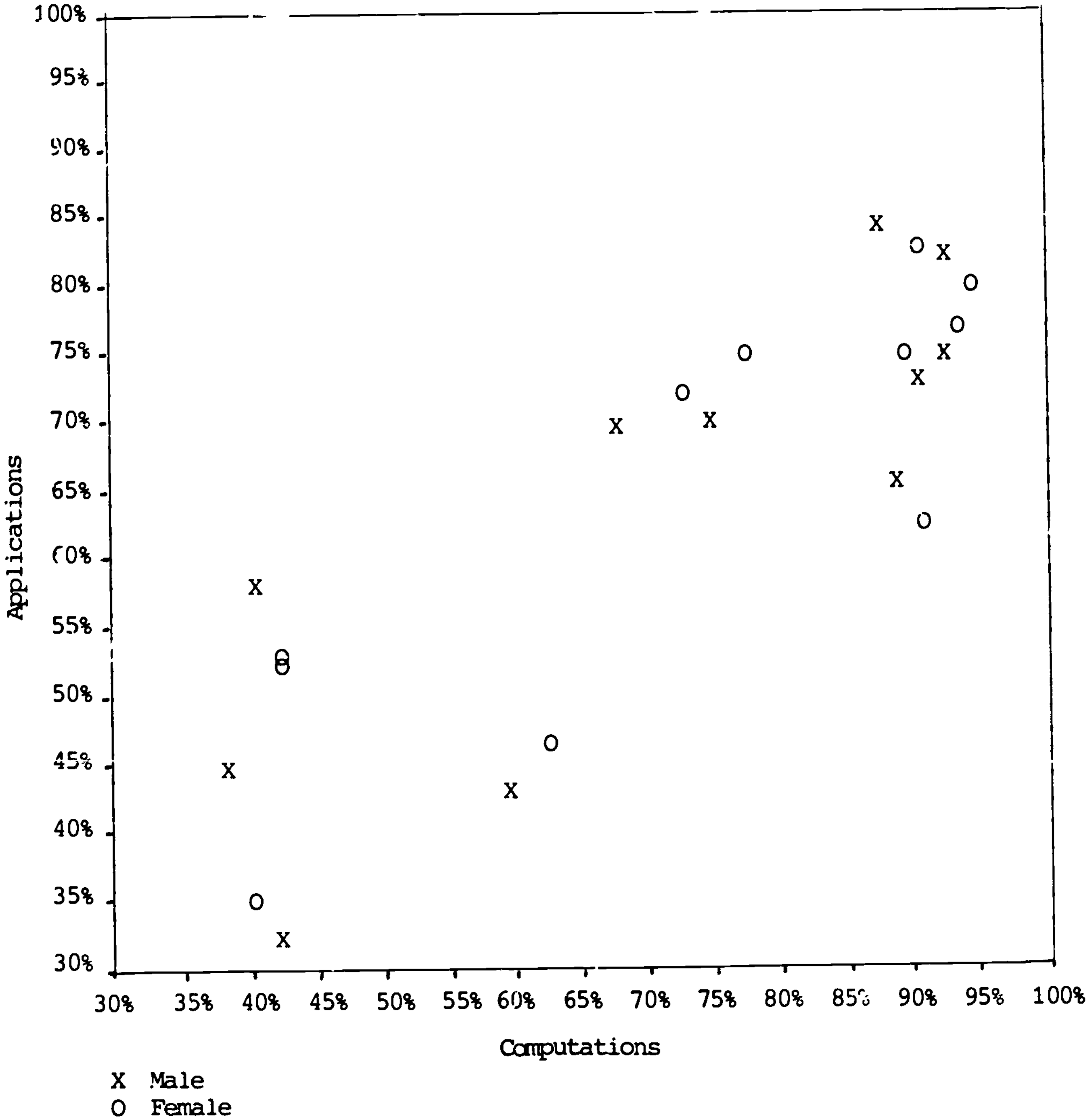


Figure 4: Performance of Boys and Firls on Matched Items of Computations and Applications: Grade 6.

2?

Analysis of Errors

Insofar as was possible, we attempted to map the distracters from the sixth-grade items into the categories defined for the third-grade items. The results are given in Table 4. Several categories were untestable at both grades, particularly distracters corresponding to erroneous rules. Only erroneous rules of subtracting the smallest from the largest value regardless of placement and carrying/borrowing errors are evaluated at both grade levels. The remaining five rules from third grade did not appear as distracters at the sixth grade. A new erroneous rule was added at the sixth grade for fraction arithmetic. This is the error of adding numerators and adding denominators when adding two fractions.

At the sixth grade, there were insufficient numbers to evaluate differences between boys and girls on errors of literal translations and on errors of interference in computing series. The first of these was marginally significant at the third grade, with girls more likely than boys to make the error, and the second revealed no difference between boys and girls. It would be useful to pursue these differences at the sixth grade as well, but the data do not allow it.

The different errors tested in this data are grouped into six categories in Table 8. The first three categories have only one error each that could be evaluated. The remaining three categories are characterized by several different errors. When these are combined within category, it is possible to evaluate whether boys or girls are more likely to make errors of erroneous rule, of association, and of attention. As mentioned above, boys are more likely to make errors from erroneous rules and girls are more likely to make errors of association. While children of both gender make errors of attention, the nature of the errors is quite distinct. Girls tend to leave out steps in multi-step calculations more often than do boys. When this error is removed from the category labeled attention, the remaining errors show significant gender difference, with boys more likely to make all the types of attentional errors than are girls ($p < .05$).

Table 8

Distracter Analysis -- Sixth Grade

TYPES OF DISTRACTERS	NO. OF ITEMS HAVING THIS DISTRACTER	NO. OF TIMES BOYS MAKE ERROR MORE THAN GIRLS	NO. OF TIMES GIRLS MAKE ERROR MORE THAN BOYS	
I. LANGUAGE ERRORS	17	10	7	
A. Opposites	17	10	7	
II. SPATIAL UNDERSTANDING	10	8	2	**
A. Spatial Reversals	10	8	2	**
III. MASTERY	66	32	34	
A. Wrong Operation	66	32	34	
IV. ASSOCIATION	39	14	25	**
A. Key Words	12	3	9	**
B. Number Patterns	27	11	16	
V. ERRONEOUS RULES	55	44	11	***
A. Small from Large	11	10	1	***
B. Carrying Errors	8	8	0	***
C. Borrowing Errors	19	14	5	***
D. Numerical Reversals	9	5	4	
E. Fraction Addition	8	7	1	***
VI. LACK OF ATTENTION	106	60	46	
A. Omit a Step	35	10	25	***
B. Careless Transcription	11	10	1	***
C. Perseverance	29	22	7	***
D. Partial Reading	19	11	8	
E. Interference: Formulas	12	7	5	

* $.10 < p < .20$
** $.05 < p < .10$
*** $p < .05$

31

Thirteen distracters are evaluated in Table 8. Of these, seven are significantly different for boys and girls beyond the .05 level of probability, and two others are marginally significant. Boys were more likely than girls to make errors caused by erroneous rules; all hypotheses tested in this category were highly significant. Girls were more likely than boys to make errors of attention and/or association, including focusing on key words, using number patterns within the problem, and omitting a needed step in the calculations. Boys also showed a propensity for errors reflecting lack of attention. In particular, they were more likely than girls to err by careless transcription of numbers and perseverance in carrying out the same procedure multiple times within a problem.

RESULTS OF THE LONGITUDINAL STUDY

As described earlier, the sixth-grade students studied here are the same students who responded to the California Assessment Test at the third grade. Their performance at the two grades is assessed by two sets of analyses: comparisons of performance at third and sixth grade on the general categories described in Table 1 and comparisons of performance on specific items of the two tests. The first set of analyses is based on the entire population of students responding at third grade and at sixth grade. The second set is based upon a subset of students whose responses at third grade could be matched uniquely with their responses at sixth grade.

Analyses Using the Entire Population

The purpose of this section is to tie together the findings of the two previous sections. In those sections, the two grades were treated separately. The objective here is to describe the continuities and discontinuities in student performance over the three year span. As before, the focus is on correct performance and on types of errors committed.

Correct Performance

Tables 2 and 6 provide information about boys' and girls' success in solving six types of problems: computations, visual items, counting/property of numbers items, word problems, geometry/measurement items, and nontraditional problems. These were discussed in the previous two sections of this report.

The range of percent correct at the third grade is 10.54 for boys and 11.28 for girls. At the sixth grade, the range for boys is 6.65 and for girls is 10.69. Clearly, there is greater similarity between boys and girls at the third than at the sixth grade.

Rank difficulty of items for boys and for girls did not change significantly from third to sixth grade. At both grades, word problems were the most difficult items for boys and for girls. Computation, visual, and counting items were consistently the easiest. There was no sex-related difference in rank order and no change over time.

Comparisons were made at both grade levels of the performance of girls and the performance of boys on each item category. Recall that at the third grade, girls performed significantly better than boys on three of the six categories: computations, nontraditional items, and counting items. Boys demonstrated no significant superiority on any category. At the sixth grade, girls continued to perform better than boys on items of computations but lost ground on four of the six categories. First, they lost the advantage demonstrated at third grade in the areas of counting and nontraditional items. There were no gender differences in performance at sixth grade for these categories. Second, boys moved from approximately equal performance with girls at third grade to superior performance in these two areas at sixth grade. These differences were statistically significant ($p < .05$).

These differences are represented graphically in Figure 5. Plotted in this figure are the differences between boys and girls on each category at each grade level. It is evident that all categories except computation evidence a shift in performance to boys' advantage from the third to the sixth grade.

Comparisons of Errors

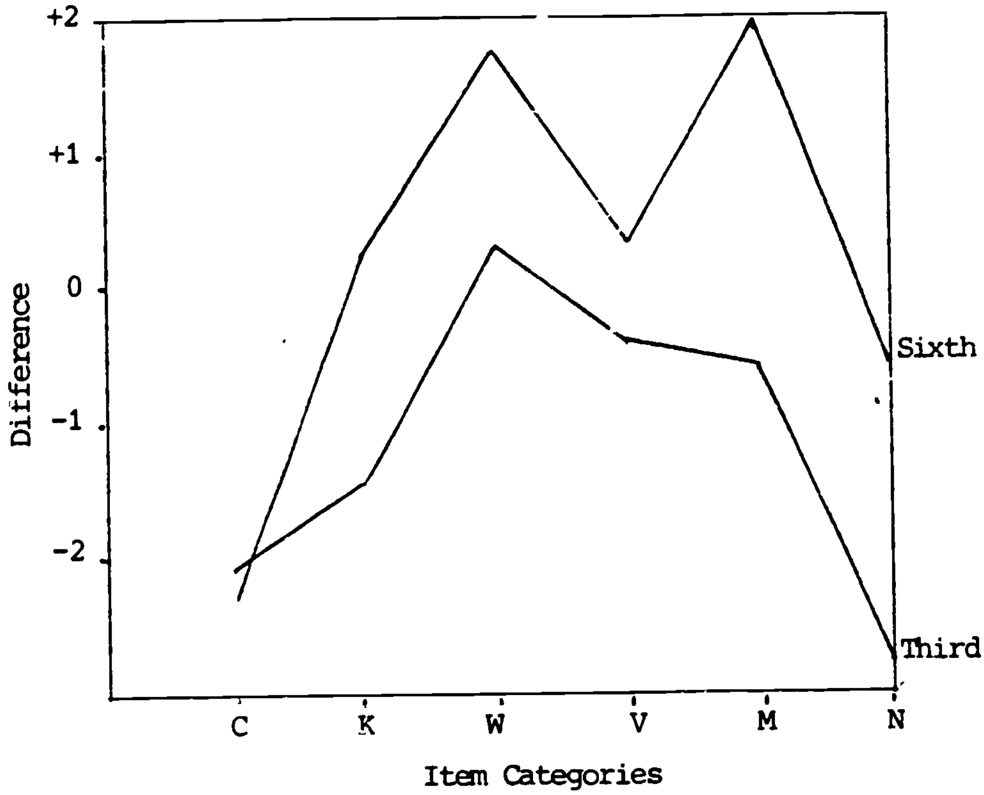
Tables 2 and 4 contain details of the comparisons between boys and girls for the selection of particular types of distracters. Most of the categories and subcategories of these two tables are identical, providing us with information about changes in distracter choices between third and sixth grades. There are three general questions to be asked:

(1) Are the sex differences observed at the third grade present to the same extent at the sixth grade?

(2) For which distracters is there increased differentiation on the basis of gender?

(3) Which distracters reflect a lessening of the differentiation observed at the third grade?

Table 9 lists the distracters and the degree of statistical significance found at each grade level. It is evident from Table 9 that boys consistently make more errors associated with erroneous rules than do girls. Six of the seven consistent findings across the grades are errors of this type.



C Computation
K Counting/Number Properties
W Word
V Visual
M Geometry/Measurement
N Nontraditional

Figure 5: Relative Change in Boys' and Girls' Performance from Third to Sixth Grade. For each item category at each grade, the average p-value for girls is subtracted from the average p-value for boys.

Table 9

Comparisons of Distracter Choices: Third and Sixth Grades

Result	Distracter Category	Comments
Significant to same degree:		Who Makes Error:
	Key Words	Girls
	Spatial Reversals	Boys
	Subtract Small	Boys
	Borrowing	Boys
	Carrying	Boys
	Careless Trans.	Boys
	Perseverance	Boys
Increased differentiation:		Who Increased No. of Errors:
	Omit a Step	Girls
Decreased differentiation:		Who Decreased No. of Errors:
	Number Patterns	Girls
	Opposites	Boys
	Numerical Reversals	Boys
No differentiation at either grade:		
	Wrong Operation	
	Partial Reading	
	Interference	

36

Comparisons Based Upon Subset of Population

For these analyses, we matched the responses of students who responded to one of the third-grade tests with their corresponding responses to one of the sixth-grade tests. We were able to isolate matching data for approximately one-third of the total population, resulting in a subset of about 100,000 students. Since there are 30 distinct forms of the third grade test and 40 forms of the sixth grade tests, we have between 80 and 90 students responding to any pair of third-sixth tests and consequently to any pair of third-sixth items.

The analyses described in this section are based upon comparisons at the item level. There are two sets of analyses. The first of these compares correct-incorrect responses at both grades upon items having matched content. The second compares selection of the same distracter on matched third-sixth grade items.

Correct Versus Incorrect Performance

The issue for these analyses is to determine whether girls and boys maintain performance in areas covered on both tests. That is, is there any gain or loss from third to sixth grade in areas that are taught at both grade levels. We focus here on content areas that are narrower than the broad categories discussed above (questions about odd and even numbers would be one such area). The comparison of interest is illustrated by a two-by-two table of correct and incorrect performance:

		Sixth-Grade Item Correct	Sixth-Grade Item Incorrect
Third-Grade Item	Correct		
	Incorrect		

Obviously, there are four cells of the table corresponding to correct performance on an item at both grades, incorrect performance at both grades, correct at third but incorrect at sixth grade, and incorrect at third but correct at sixth. For every matched pair of items (and for every item repeated on both sixth and third grade tests), a simple contingency table can be constructed for boys and girls. Within a particular content area, one can aggregate frequencies from all matched items. The test of interest is a chi-square test of distribution: testing whether the distribution of boys over the four cells is similar to the distribution of girls.

The narrow content areas investigated here are given in Table 10, together with a brief description of each.

Table 10

Specific Content Areas Present on Third-Grade and Sixth-Grade Tests

Content Area	Brief Description
1. Place Value I	Identify the digit in a specified place value.
2. Place Value II	Recognize the place value of a specified digit.
3. Place Value: III	Write number in expanded notation using place value.
4. Odd/Even	Identify odd and even integers.
5. Identify Question	Recognize restatement of the question asked in word problem.
6. Fraction Shaded	Identify the fraction shaded in a figure.
7. Identify Function Rule	Recognize relation between X and Y, given multiple numerical examples of each.
8. Find Missing Number	Find value of X, given Y and multiple instances showing relation between X and Y.
9. Geometry: Parallel Lines	Identify parallel lines.
10. Geometry: Line Segments	Identify line segments.
11. Geometry: Perimenter	Find the perimeter of a figure.
12. Geometry: Graphs	Identify coordinates in a graph.
13. Fill in the Box	From a worked-out computation, find missing value that is represented by a box.
14. Words to Numbers	Translate written numerical statement to numbers.
15. Equation to Problem	Given numerical equation, identify appropriate word problem that matches it.
16. Problem to Equation	Given word problem, identify corresponding mathematical expression or equation.
17. Word Problems -- Change	Permanent alteration in some set.
18. Word Problems -- Combine	Two distinct sets are joined into a conceptual superset.
19. Word Problems -- Compare	Contrast the difference between two quantities.
20. Word Problems -- Vary	Direct variation of one quantity with a second one.
21. Word Problems -- Transform	Expressing a quantity in a different scale/dimension.

Two questions were asked of each content area. First, did boys and girls differ in their progressions in these areas? This corresponds to the comparisons of percent correct of boys and of girls at the third grade with percent correct of boys and girls at the sixth grade. The second questions is whether the distributions of boys and girls differ in the two-by-two contingency tables.

The answers to the first question are given in Table 11. The performance of girls relative to that of boys drops in 9 of 21 areas (43 percent). The performance of boys relative to that of girls drops in 5 of the areas (24 percent). The difference between boys' and girls' performance remains stable from third to sixth grade for the remaining 7 areas (33 percent).

Most of the areas in which girls appear to lose ground are from the counting/property of numbers category. As pointed out previously, this is an area in which girls performed significantly better than boys at the third grade but not at the sixth. The analyses of Table 11 illustrate the particular difficulties that were experienced by girls.

There is no clear pattern to the relative loss by boys. Two of the five areas are geometric, and two are types of word problems. Boys scored significantly higher on both general categories than did girls at the sixth grade.

Rather than look at these differences as relative losses, we can view them as relative gains. Thus, boys made relative gains in the area of counting/property of numbers from third to sixth grade. Similarly, girls made relative gains in two areas each of geometry and word problems.

Knowing whether one sex makes relative gains does not answer the question of whether girls and boys are responding in approximately equal ways to the pairs of items. In particular, it does not provide any information about the distribution of student responses over the four cells. Table 12 contains the results of chi-square tests of distributions for the 21 content areas.

Table 11

Comparison of the Relative Gains by Boys and Girls from Third to Sixth Grade

GIRLS LOSE GROUND:	No difference at third grade; boys surpass girls by at least 5% at sixth.	1. Place Value I 2. Place Value II 3. Place Value III 10. Geometry: Line Segment 13. Fill in the Box
	Girls surpass boys by at least 5% at third grade; no difference at sixth.	4. Odd/Even 5. Identify Question 6. Fraction Shaded 15. Equation to Problem
BOYS LOSE GROUND:	No difference at third grade; girls surpass boys by at least 5% at sixth.	7. Identify Function Rule 19. Word Problems -- Compare 21. Word Problems -- Transform
	Boys surpass girls by at least 5% at third grade; no difference at sixth.	11. Geometry: Perimeter 12. Geometry: Coordinate Graphs
NO CHANGE:	Boys and girls approximately equal at both grades.	8. Find Missing Number 9. Geometry: Parallel Lines 17. Word Problems -- Change 18. Word Problems -- Combine 20. Word Problems -- Vary
	Boys surpass girls by more than 5% at both grades.	14. Words to Numbers 16. Problem to Equation

Table 12

Comparisons of Boys' and Girls' Responses to Matched Pairs of Items

	Content Area	Chi-Square	Level of Significance
1.	Place Value I	3.03	
2.	Place Value II	4.17	
3.	Place Value: III	4.68	
4.	Odd/Even	18.36	***
5.	Identify Question	16.05	***
6.	Fraction Shaded	27.24	***
7.	Identify Function Rule	12.79	***
8.	Find Missing Number	4.73	
9.	Geometry: Parallel Lines	8.07	**
10.	Geometry: Line Segment	2.97	
11.	Geometry: Perimeter	7.38	*
12.	Geometry: Coordinate Graphs	7.34	*
13.	Fill in the Box	7.07	*
14.	Words to Numbers	14.45	***
15.	Equation to Problem	5.69	
16.	Problem to Equation	15.27	***
17.	Word Problems -- Change	10.11	**
18.	Word Problems -- Combine	4.57	
19.	Word Problems -- Compare	7.85	**
20.	Word Problems -- Vary	77.38	***
21.	Word Problems -- Transform	14.22	***

* $p < .10$ Marginally Significant
** $p < .05$ Significant
*** $p < .01$ Significant

Eleven of the 21 tests yield results that exceed conventional levels of statistical significance. An additional 3 are marginally significant The important questions in these tests are whether girls and boys are equally likely to miss both items and whether they are equally likely to solve correctly the third grade item and err on the sixth grade one. A large number of students were unable to solve either the third or sixth grade item in all categories. Over ten percent of all students solving items in 13 of the 21 categories missed both items. Girls were more likely than boys to exhibit this pattern of response. In particular, on two categories, identifying the function rule (7) and solving vary word problems (20), over 20 percent of the girls failed to solve the items at either grade.

Six of the significant findings of Table 12 are in areas related to solving word problems: identifying a restatement of the question asked in a word problem (5), recognizing the equation or expression that corresponds to a word problem (16), and solving four categories of word problems. A closer examination of the distributions of boys' and girls' responses in these categories indicates that the major differences are in the correct-correct and correct-incorrect cells. Girls are more likely to be correct on both pairs of items than are boys for four of the areas. Boys are more likely than girls to be correct on the third-grade item and incorrect on the same type of sixth-grade item. This pattern was observed in five of the six areas.

Persistence in Making the Same Error

One goal of the research described in this report was to examine the extent to which children continue to make the same types or errors over time. This issue was addressed by taking the five categories of word problems described in Table 10 and examining student performance on those which had parallel distracters. For example, there were change word problems at both grades that required addition for the correct solution and offered the distracter of subtraction.

Analyses similar to those described above were carried out. For each set of matching items, the following two-by-two contingency table could be developed:

		Sixth-Grade Item	
		Correct	Distracter
Third-Grade Item	Correct		
	Distracter		

There are two questions of interest. First, are there significant differences in the distributions of boys' and girls' responses to these items? Second, are boys or girls more likely to show persistence in making the same error?

Fourteen sets of items were identified. Each set allowed the contrast of correct performance with the same distracter. A large majority of the third-grade items were simple word problems involving a single arithmetic operation. These items were matched with a similar set of sixth-grade items also requiring only one computation. In all cases, the matching distracters corresponded to computation using the same incorrect arithmetic operation.

Of the 14 comparisons, 4 were statistically significant, with a probability level smaller than .05, and 3 were marginally significant, having a probability less than .10. The remaining 7 tests revealed no sex differences.

Only compare items were nonsignificant. There were significant differences in performance on the remaining four types of items.

4ɜ

SUMMARY

It is clear from the preceding analyses that boys and girls differ in performance at both third and sixth grades. Girls appear to be stronger than boys in mathematics at the third grade. This is evidenced in Table 2 and the analyses pertaining to that table. Girls have higher probabilities of successful performance on 5 of the 6 categories of items. Three of these differences are statistically significant. This finding has been observed by the California Assessment Program in other cohorts of students as well (CAP, 1982).

These findings contradict other research on gender-related differences in mathematics performance (e.g., Leder, 1982; Benbow & Stanley, 1982). In very few instances have girls been reported to have higher achievement than boys. There are several points to be made in this regard. First, we can be very sure of these results. The data are not a sample from a specific population; the entire population of third-grade students in California was examined. The test itself is a broad-range instrument containing 360 items. Thus, the present results cannot be dismissed as artifacts of sampling either from the population of students or from a limited range of items.

We gain some information about why girls have higher performance than boys from the analyses of errors. Most of the statistically significant findings relate to boys' tendencies to apply erroneous arithmetic rules to mathematics problems (see Table 4). We hypothesize that girls are more likely to develop and use the rules of mathematical computation. Confirmation of this hypothesis requires additional research on children's abilities to identify, formulate and discriminate among different rules or algorithms. There also seems to be an element of attention to detail reflected in the analyses, particularly in the tendency of boys to make careless errors of spatial or numerical reversals. Again, further research is required to determine whether this difference is atttentional, developmental, or gender-related.

In the sixth-grade results, we observe more traditional results. Boys seem to do better than girls. Girls have higher probabilities of success on only two of the six categories of items. Previous research on an earlier version of the California Assessment Test indicated that girls were more likely than boys to solve computational items correctly and boys were more likely than girls to solve word problems correctly (Marshall, 1984). These results were replicated here. However, two additional findings of the present research complicate a simple interpretation of the computation/word problem results. First, boys and girls demonstrate approximately equal understanding of counting principles and properties of numbers. Thus, girls may be more likely than boys to carry out computations correctly, but they are not more likely to demonstrate understanding of the computations. This suggests again a dependency upon rules or algorithms for computations with or without a clear understanding of what the rules mean.

The second additional finding is that girls are more likely than boys to solve nontraditional word problem correctly and less likely to solve traditional ones correctly. The primary distinction between these two types of items is that in the former students are not asked to reach a numerical solution. They are expected to analyze a problem, interpret intermediate steps, identify operations that will be required, and so forth. The fact that girls consistently have higher performance than boys on problems of this type (see CAP, 1982) suggests that girls do indeed have the capability of understanding what is happening within a word problem. Why, then, do they perform more poorly than boys on word problems? There are several possible explanations. One is that girls develop different reading skills for mathematics problems. They may engage in spot reading or in searching for selected words in the text. When asked to produce a novel response as in the nontraditional items, they may change their reading styles.

A second reason for the difference in performance is the rule argument presented above. Using rules is essential for solving computations correctly. It may be natural for girls to expect that rules also govern word problems and to develop rules that can be applied to such problems. Certainly at the very simplest level, there are a few rules that may be invoked, such as the word "altogether" generally means that addition will be required. The problem with this strategy is that it cannot generalize to complex problems requiring several arithmetic computations.

A disturbing result of this study is the comparison of boys' and girls' responses to third and sixth grade items of the same type (see Tables 10 and 11). It is here that we see more clearly what is happening in mathematical skill development between these two grades. Girls either lose ground or fail to maintain equal performance with boys on approximately one-half of the subcategories studied. This is roughly twice as many categories as those showing declining performance by boys.

These results suggest that teachers may need to address specific deficits, particularly in the underlying principles and concepts of mathematics. It may be that boys and girls require additional instruction or elaboration in different areas. In particular, it may be necessary to provide instruction about how to read "mathematically". In traditional word problems, many inferences must be made. At this point, we have no information about gender-related differences in drawing such inferences. We do know that word problems are very difficult for both boys and girls (from the rank orders of Tables 2 and 6), and we also know that they are significantly more difficult for girls than for boys. The next step must be to evaluate whether differences in the ability to read mathematically can account for the results observed here. This should be a fruitful area of research and may help us understand better how students solve mathematics problems.

REFERENCES

Benbow, C. & Stanley, J. (1982). Consequences in high school and college of sex differences in mathematical reasoning ability: A longitudinal perspective. American Educational Research Journal, 19, 598-622.

California Assessment Program. (1981). Student Achievement in California Schools, 1980-81 Annual Report. Sacramento: California State Department of Education.

California Assessment Program. (1982). Student Achievement in California Schools, 1981-82 Annual Report. Sacramento: California State Department of Education.

Leder, G. C. (1982). Mathematics achievement and fear of success. Journal for Research in Mathematics Education, 13, 124-135.

Marshall, S. P. (1984). Sex differences in children's mathematics achievement: Solving computations and story problems. Journal of Educational Psychology, 76, 194-204.

Marshall, S. P. (1983). Sex differences in mathematical errors: An analysis of distracter choices. Journal for Research in Mathematics Education, 14, 325-336.

Marshall, S. P. (1982). Sex differences in solving story problems: A study of strategies and cognitive processes. Final Report, NIE-G-80-0095, The National Institute of Education.

APPENDIX A

Sample Items From California Assessment Program Tests
From Third and Sixth Grades

PROBLEM TYPE	GRADE	EXAMPLE
Computation	Third	740 -672 ----- (*) 68 ( ) 78 ( ) 132 ( ) 1412
	Sixth	1/5 + 3/4 = ( ) 4/9 (*) 19/20 ( ) 4/20 ( ) 3/20
Counting	Third	345 = ( ) 3 + 4 + 5 ( ) 400 + 30 + 5 ( ) 400 + 50 + 3 (*) 300 + 40 + 5
	Sixth	To find the <u>difference</u> between 83 and 18, you: ( ) add (*) subtract ( ) multiply ( ) divide

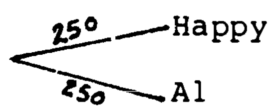
---

Word Problems	Third	Ron had 7 peanuts. Sue had 2 times as many peanuts as did Ron. How many peanuts did Sue have? ( ) 21 ( ) 16 (*) 14 ( ) 9
	Sixth	It is 1.3 kilometers from Sharon's house to school. She rides her bicycle to and from school every day. How far does she ride in 5 days? ( ) 6.3 kilometers ( ) 6.5 kilometers ( ) 10 kilometers (*) 13 kilometers
Nontraditional Problems	Third	Kim had 4 apples. She ate 3. How many were left? .... Which question is asked? ( ) Did Kim have 4 apples? (*) How many were left? ( ) How many apples did Kim eat? ( ) Did Kim have 3 apples left?
	Sixth	The 130 students from Marie Curie School are going on a picnic in Carson Park. Carson Park is 12 miles from the school. Each bus holds 50 passengers. How many buses are needed? Which numbers are needed to solve this problem? ( ) 130 and 12 (*) 130 and 50 ( ) 12 and 50 ( ) 130, 12, and 50

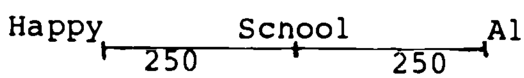
Geometry /Measurement	Third	This shape is: ( ) a circle ( ) a square (*) a triangle ( ) a rectangle
	Sixth	A hand is used to measure the height of a horse. The hand is 4 inches long. How tall is a horse that measures 15 1/2 hands? ( ) 15 1/2 inches (*) about 5 feet ( ) about 62 feet ( ) 15 1/2 feet

Visual	Third	Jenny was saving pennies. She put them in bags of 10's and 100's. How many pennies does Jenny have? ( ) 423 ( ) 342 (*) 324 ( ) 304
	Sixth	Happy and Al lived the same distance from the school but in opposite directions. They found that they lived 500 meters apart. Which drawing shows this?

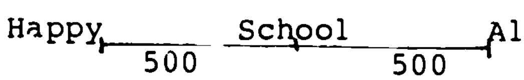
( ) School



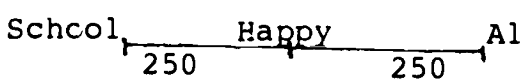
(*)



( )



( )



These items are reproduced here with the permission of the California Assessment Program.