DOCUMENT RESUME

ED 222 557 TM 820 711

AUTHOR Harnisch, Delwyn L.; Linn, Robert L.
TITLE Identification of Aberrant Response Patterns. Final Report.
INSTITUTION Education Commission of the States, Denver, Colo. National Assessment of Educational Progress.; Illinois Univ., Champaign.
SPONS AGENCY National Inst. of Education (ED), Washington, DC.
PUB DATE [81]
GRANT NIE-G-80-0003
NOTE 120p.; For related documents, see TM 820 707-712 and TM 820 716.

EDRS PRICE MF01/PC05 Plus Postage.
DESCRIPTORS Educational Assessment; *Educational Diagnosis; *Error Patterns; Item Analysis; Mathematics Achievement; National Surveys; Response Style (Tests); *Scores; Secondary Education; *Student Characteristics; Student Evaluation; *Test Interpretation
IDENTIFIERS Caution Index; National Assessment of Educational Progress; *NIE ECS NAEP Item Development Project; Response Patterns; Second Mathematics Assessment (1978)

ABSTRACT

Techniques to identify the degree to which a response pattern is unusual and the pattern's relationship to examinee background are presented. Sato's caution index and modified caution index use clusters of items for comparisons of observed performance outcomes and do not require the use of item response theory. Correlation between the indices was nearly one for the indices computed for the students and the problems. The algebraic formulations of the indices are discussed with their empirical interrelationships and relationship with total number-right scores for data collected by the National Assessment of Educational Progress (NAEP) from 13-year-olds in the 1978 Mathematics Assessment. High Caution indices of unusual response patterns for individual test examinees are discussed. The indices may be useful in interpreting a total test score and may identify particular strengths and gaps in examinee knowledge. The distributional and relational properties of the indices with total scores, indentification of student background characteristics with unusual response patterns, and school characteristics associated with aberrant response patterns are shown. The variables associated with the modified caution index and comparisons of subgroup caution index values are discussed. Primary type of information provided by the report: Results (Secondary Analysis). (CM)

FINAL REPORT

Grant Title: Identification of Aberrant Response Patterns

National Institute of Education

Grant Number: NIE-G-80-0003

Investigators: Dr. Delwyn L. Harnisch
Assistant Professor of Educational Psychology
123 Institute for Child Behavior and Development
University of Illinois, Urbana-Champaign
51 Gerty Drive
Champaign, Illinois 61820
(217) 333-4123

Dr. Robert L. Linn
Professor of Educational Psychology and of Psychology
210 Education Building
University of Illinois, Urbana-Champaign
1310 South Sixth Street
Champaign, Illinois 61820
(217) 333-2245

The work upon which this report is based was performed pursuant to ECS Contract No. 02-81-20318 which was issued under a jointly sponsored project of the Education Commission of the States (ECS) and the National Institute of Education (NIE). It does not, however, necessarily reflect the views of ECS or NIE.

Grant Period: January 1, 1981 - December 31, 1981

Table of Contents

List of Exhibits

List of Tables

List of Tables (continued)

List of Figures

List of Figures (continued)

List of Appendices

INTRODUCTION

Millions of achievement tests are administered each year to students in this country. The types of tests are quite varied as are the purposes of the testing. Tests are often used, as in the case of the National Assessment of Educational Progress and several statewide testing programs, to provide a general assessment of large aggregates of students. Survey tests are often used to monitor achievement of students at the school or district level. Individual student achievement is also assessed for many purposes. To name a few, test results are used to report achievement to parents, for diagnosis of learning difficulties, for grade-to-grade promotion and for the award of high school diplomas. Mandated test use in program evaluations and the selection of students for special programs accounts for another large segment of test use.

In most uses of achievement tests, whether with individual pupils or with aggregates, the primary focus is on a few summary scores that are based on the number of items answered correctly. The focus on the number of correct answers or some transformation of it (e.g., a grade-equivalent score) is natural. Certainly, the overall level of performance in a content area such as arithmetic is a major consideration. Furthermore, the interrelatedness of subsets of achievement test items (e.g., addition and subtraction items) makes it difficult to abstract reliable information about special strengths and weaknesses for individuals or for groups of students in addition to a global summary score.

Nonetheless, it is clear that the same number-right score can be obtained in many different ways. Even on a short test of just five items, a score of three right answers can be obtained by ten different combinations of correct and incorrect responses. The number of possible combinations of right and wrong answers expands rapidly with increases in the

test length. On a twenty-item test, for example, there are 184,756 different response patterns that yield the same number right score of ten.

Although the 184,756 distinct patterns of response can be identified in the above example, it obviously is not feasible to provide different interpretations for each unique pattern. Responses to individual items are too unreliable. Many of the possible patterns will not be encountered in practice and the frequency of occurrence of any particular pattern will usually be too small to support any clear generalizations beyond those that can be made from the simple number-right score.

The difficulties in using response patterns to obtain diagnostic information not contained in the number-right score explain the relatively limited use that has been made of such information. Recently, however, some new and potentially more powerful techniques have been developed for identifying atypical response patterns. The introduction of these techniques has led to a renewed interest in using information contained in response patterns to identify individuals with unusual patterns (see, for example, Harnisch & Linn, 1981; Levine & Rubin, 1979; Sato, 1975, 1981; Tatsuoka & Tatsuoka, 1980; van der Flier, 1977).

The identification of unusual response patterns for individuals has several possible uses. The most obvious use is simply to identify students for whom special caution is needed in interpreting their total correct scores. One of the purposes of this study is to explore the use of an index of the degree to which an individual's response pattern is unusual and whether this index is related to background characteristics of the individual test takers.

The identification of persons for whom special caution is required in interpreting total scores is important. Potentially even more important, however, is the possible diagnostic value of the response pattern.

For example, the content characteristics of items that contribute to a high caution index for a student may identify particular strengths or gaps in the student's knowledge.

The role of specific test content often receives relatively little attention when the ranking of students on a global dimension is emphasized. For some purposes, it is useful, for example, to know that students in an instructional program have an average score well below the national mean of a fourth-grade mathematics test. But, by itself, this statement is uninformative about the types of arithmetic skills the students have. As shown by Porter, Schmidt, Floden, and Freeman (1978), there is wide variation in content coverage of the fourth-grade mathematics tests from the four most widely used achievement test batteries. Furthermore, the level of performance can be expected to be influenced by the degree of overlap between the instructional content and the content of the test (e.g., Madaus, Kellaghan, Rakow, & King, 1979). Indeed, the specific match between the format of instructional exercises and the format of test questions can have a substantial impact on test scores (Alderman, Swinton, & Braswell, 1979; House, Glass, McClean, & Walker, 1978).

To the degree that content coverage is important and that schools vary in their content coverage, then schools would be expected to vary not only in their overall performance as measured by the total test score, but in the relative difficulty of particular subsets of items. The second major purpose of this study was to determine if schools could be reliably distinguished in terms of consistent response patterns that deviated from the national norm. The demonstration of systematic between school differences in response patterns and the association of these differences

with differences in instructional practices has potentially important implications for test use in evaluation studies. The match between tests and content coverage is an important consideration in evaluation studies, but one that has frequently been given little attention. The indices that were used in this study provide a potential means of identifying situations where comparisons between groups solely on the basis of total scores may be misleading. The clusters of items that contribute to high caution indices provides alternate comparisons, and possibly more meaningful interpretations of observed performance outcomes.

REVIEW OF RESEARCH

Two major types of appropriateness indices based upon the pattern of item responses to individual items have been defined. First, there are the indices which are based upon item response theory (IRT) as described by Levine and Rubin (1979) and modifications of these indices as suggested by Drasgow (1978). The χ^2 test of person fit that is sometimes used with applications of the Rasch model (e.g., Wright, 1977) is another example of an IRT based index. Second, there are the indices which are based directly uppon the pattern of right and wrong answers, such as the "caution" index proposed by Sato (1975), the modified caution index by Harnisch and Linn (1981), the U' index by van der Flier (1977), the personal biserial by Donlon and Fischer (1968), the norm-conformity index by Tatsuoka and Tatsuoka (1980), and the agreement and disagreement indices discussed by Kane and Brennan (1980).

The focus of this final report is on two activities conducted with the latter type of indices, Sato's caution index and the modified caution index, which do not require the use of IRT. Our primary purposes are (1) to evaluate the distributional and relational properties of the caution

indices with the total test score, (2) to identify student background characteristics that are associated with unusual response patterns (i.e., with high values of C_i, or C_i*, and (3) to identify school characteristics that are associated with high incidence of aberrant response patterns. We will review briefly the caution indices and their algebraic formulation. We will then discuss the empirical interrelationship among the caution indices and their relationships with total number-right scores for data collected by National Assessment of Educational Progress (NAEP) from 13-year-olds in 1978 on two mathematics test packages. We will then discuss the student and school characteristics associated with high modified caution indices. Finally, we will examine the school and census division differences on the modified caution index, and content differences in items which contribute to school-to-school differences in the index.

DEFINITION AND COMPARISON OF INDICES

For purposes of defining and comparing indices of the degree to which an individual's response pattern is unusual it is convenient to start with a consideration of a matrix of zeros and ones. A row of the matrix is associated with each examinee and a column with each item. Ones are recorded for correct responses and zeros for incorrect response. Rows and columns of the data matrix are permuted so that the items (columns) are arranged, from left to right in ascending order of difficulty, and examinees (rows) are arranged, from top to bottom in descending order of total number of correct answers. The resulting matrix has been called an "S-P" Table (student-problem table) by Sato (1975). (See Tatsuoka [Note 1] for a description in English.)

If the items formed a perfect Guttman Scale (Guttman, 1941), the S-P Table would consist of a section with all ones in the upper left-hand

corner and all zeros in the lower-right-hand corner. A single step-like boundary line would separate the ones and the zeroes. In other words, anyone who responded correctly to a difficult item would also answer all easier items correctly. There would be no unusual response patterns in the sense that is used for the indices described in this manuscript because everyone who had a given total score would have the same pattern of responses. Of course, with responses to achievement test items, perfect Guttman scales cannot be expected. Consequently, the S-P Table will be characterized by a predominance of ones in the upper-left hand corner and zeroes in the lower right-hand corner, but there will be many exceptions to the pattern, i.e., ones in the region where mostly zeroes are found and vice versa.

A small hypothetical example of an S-P Table with 18 examinees and 5 items is shown in Table 1. The solid and dashed lines in Table 1 are known as the S-curve and P-curve respectively. The S-curve (solid line) is obtained by drawing a vertical line for each row that has $n_{i\cdot}$ items (columns) to the left of it where $n_{i\cdot}$ is the total number of correct responses for the i^{th} examinee. The P-curve (dashed line) is obtained by drawing a horizontal line in each column such that there are $n_{\cdot j}$ examinees who answer item j correctly.

- - - - - - - - - - - - - -

Insert Table 1 about here

- - - - - - - - - - - - - -

For an ideal, or Guttman-scalable S-P Table, the S- and P-curves would coincide. The degree of divergence indicates heterogeneity, degree of convergence indicates homogeneity of the response patterns. Sato (1975) has developed an index based on the area between the S- and P-curves which is

potentially useful in evaluating the homogeneity of the test (see Tatsuoka, Note 3). Of greater interest for our present purposes, however, is Sato's "caution" index.

Sato's caution index, C_i for the $i^{\underline{th}}$ examinee, may bd defined as follows:

$$C_i = \frac{\sum_{j=1}^{n_{i\cdot}} (1 - u_{ij})\, n_{\cdot j} - \sum_{j=n_{i\cdot}+1}^{J} u_{ij}\, n_{\cdot j}}{\sum_{j=1}^{n_{i\cdot}} n_{\cdot j} - j_{i\cdot} \frac{\sum_{j=1}^{J} n_{\cdot j}}{J}}, \qquad (1)$$

where $i = 1, 2, \ldots I$, indexes the examinee,

$j = 1, 2, \ldots J$, indexes the item,

$u_{ij} = 1$ if examinee i answers item j incorrectly,

$u_{ij} = 0$ if examinee i answers item j incorrectly,

$n_{i\cdot}$ = total correct for the $i^{\underline{the}}$ examinee, and

n_{ij} = total number of correct responses to the $j^{\underline{th}}$ item.

A parallel index for the $j^{\underline{th}}$ item may be defined by simply reversing the roles of i and j in the above equation, but only the person index will be considered in the present paper. Values of C_i, and for all of the indices described below, are listed in Table 1.

The name of the index comes from the notion that a large value is associated with examinees that have unusual response patterns. It denotes that some caution may be needed in interpreting a total correct score for an examinee. An unusual response pattern may result from guessing, carelessness, high anxiety, an unusual instructional history or other experiential background, a localized misunderstanding that influences responses to

a subset of items, or copying a neighbor's answers to certain questions. The key point is that the caution index provides information about an examinee that is not contained in the total score. A large value of the caution index raises doubts about the validity of the usual interpretation of the total score for an individual.

A modified form of Sato's caution index, C_i^*, was introduced to yield a lower bound of 0 and an upper bound of 1. This modified caution index, C_i^*, for the $i^{\underline{th}}$ examinee may be defined as follows:

$$C_i^* = \frac{\sum_{j=1}^{n_{i\cdot}} (1 - u_{ij}) n_{\cdot j} - \sum_{j=n_{i\cdot}+1}^{J} u_{ij}\, n_{\cdot j}}{\sum_{j=1}^{n_{i\cdot}} n_{\cdot j} - \sum_{j=J+1-n_{i\cdot}}^{J} n_{\cdot j}} \qquad (2)$$

By being bound between 0 and 1, the modified caution index eliminates the extreme scores that are sometimes obtained on the caution index, especially in cases where a very high scoring examinee misses a single very easy item. This is seen as a potential advantage of C_i^*.

Sato's caution index and the modified caution index have been briefly described. Both indices provide an indication of the degree to which an individual's response pattern departs from a norm.

DATA SOURCE AND PROCEDURE

The data used in this investigation came from Booklet 4 and 5 given to thirteen-year olds during the 1977-1978 NAEP mathematics survey. Booklets 4 and 5 were selected because they contained the affective domain items related to attitudes toward mathematics along with the cognitive items related to math achievement. Public use data files were purchased

from the Education Commission of the States (ECS) (Note 2). Individual responses of 9-, 13-, 17-year olds to background, affective domain, and cognitive items are recorded for numerous booklets used at each respective age in the national survey.

Sampling

A deeply stratified, multi-stage sampling design with oversampling of low-income and rural areas was used by NAEP (National Center for Educational Statistics, 1978; Moore, Chromy, & Rogers, 1974). In the first stage, the United States is divided into geographical units, including Standard Metropolitan Statistical Areas and counties or groups of contiguous counties. These are stratified according to region and size of community. In the second stage, schools are randomly sampled within selected first-stage geographical units. Finally, from 10 to 35 age-eligible students are randomly selected within schools for testing.

The NAEP test materials are organized into booklets suitable for administration in a single sitting. Each booklet or "test package" contains a collection of items taken by a selected student. In the 1978 mathematics assessment, 11 separate booklets were each administered to large national samples of 13-year-old students.

Instruments

Booklet 4. Affective items in Booklet 4 asked students questions about how much they liked mathematics and how useful they believed it was. The list of affective items is given in Exhibit 1, used a Likert scale and were scored 1 to 5. Affective items were recoded so that a value of 5 would indicate a greater liking of math or a greater usefulness of mathematics. One additional non-achievement test item scored "yes," "no" or "I don't know" was used from this booklet. The question asked of students was: Do you or your family own a hand calculator?

Items from the principal's questionnaire included in this study were the school size, type of school (public versus private Catholic), percent of students black, number of students qualified to receive Title I and the percentage of students attending school whose parents are professional personnel, skilled workers, unskilled workers, farm workers, not employed, and on welfare.

Cognitive items from Booklet 4 were categorized in terms of their content and format into four categories: arithmetic, geometry, tables and graphs, and miscellaneous. The listing of the cognitive item numbers into their respective categories is given in Appendix A.

Booklet 5. Affective items in Booklet 5 consisted of statements to which respondents indicated the easiness, importance and the degree to which they liked mathematics. Exhibit 2 provides a listing of affective domain items given in Booklet 5 which used a Likert scale and were scored 1 to 5. The items were recoded so that a high value would indicate a greater amount of the domain being tapped. A similar additional affective item was used from Booklet 5 as well as the items from the principal's questionnaire that were used from Booklet 4.

The cognitive items from Booklet 5 were similarly categorized into the same four categories as Booklet 4. A listing of the cognitive item numbers and respective item category is given in Appendix B.

Reliability of Affective Domain Scales

The affective items listed in Exhibit 1 were classified into the Like Mathematics scale and the Usefulness of Mathematics scale. Reliability analyses of these scales were conducted yielding a Cronbach coefficient alpha of .49 for the Like scale and an alpha of .67 for the Usefulness scale.

The affective items listed in Exhibit 2 were classified into either the Importance, Easiness, and Like scales. Reliability analyses of these scales were conducted yielding a Cronbach alpha of .71 for the Importance scale, .66 for the Easiness scale and .69 for the Like scale.

Each of these scales along with the student backgound and school variables were used for the correlational analysis with the individual caution indices computed on all cognitive items, and on subsets of arithmetic and geometry items.

Data for Secondary Analysis

The data tape provided by the NAEP contains information on 2,437 and 2,462 13-year-olds responding to questions on Booklet 4 and Booklet 5, respectively. Each sample is composed of an almost equal number of boys and girls, 50.8% boys in Booklet 4 and 48.5% boys in Booklet 5. Most students were attending a public school (87.2%--Booklet 4; 89.9%--Booklet 5) as compared to a private school. The majority of the students are white while blacks and other racial groups are represented. Most parents of students in each sample were blue-collar workers, clerical or skills workers, somewhat fewer were classified as professional, managerial or farm workers. For each of the booklet samples, parents of approximately 12% of the sample were on welfare while approximately another 7% were not regularly employed.

Data Analysis

Computer programs were written in FORTRAN IV to compute the individual and item caution indices. Statistical programs from Statistical Package for the Social Sciences (SPSS) (Nie et al., 1975), Statistical Analysis System (SAS) (SAS User's Guide, 1979 Edition), and Finn's MULTIVARIANCE were used to answer the research questions.

The NAEP sampling procedures require that individual weighting procedures be used to reflect the representative target population. In order to obtain unbiased estimates of population means we took the individual weight value supplied by NAEP on the tape file and created a weighted data file with observations being weighted in inverse proportion to their probabilities of being included in the sample. All correlational analyses were based on the weighted data file for the sampling design, and therefore the results apply to a representative population of students in the United States (Moore et al., 1974). For school analyses, a weighted procedure was followed as outlined by David Wright (1981, p. 30) which would assure the representativeness of our nations' schools.

RESULTS AND DISCUSSION

Booklet 4

A stem-and-leaf display of the percentage correct values on the 68 cognitive items used in Booklet 4 is given in Table 2. The numbers in the first column, which is labeled "stem" list the ten digits for the percent correct. The numbers to the right provide the units digit and two decimal places for individual items. For example, the easiest item was answered correctly by 97.17% of the test takers. This is denoted in the first row of Table 2 by the stem of 90 and a leaf of 7.17. The frequency of 1 indicates that only one item was answered correctly by more than 95% of the test takers (the second row lists percent correct between 90 and 95%).

As can be seen in Table 2, the item difficulties ranged in percentage correct from 2 to 97%. There are a large number of relatively difficult items. Thirty one of the 68 items, for example, are answered correctly by fewer than 40% of the test takers. Appendix A provides for each item

the percent correct, the point-biserial, Sato's caution index, and the modified caution index. Two items were identified with modified caution indices over .5 suggesting that an unexpected large number of low performing students are answering these items correctly. The means, standard deviations, and correlations among the four-item statistics are given in Appendix B. A high negative correlation was found between the point-biserial and each of the caution indices. The extremely high correlations among the caution indices for the items (.99) is not surprising. The two indices yield nearly identical information about the items.

- - - - - - - - - - - - -

Insert Table 2 about here

- - - - - - - - - - - - -

The intercorrelations among Sato's caution index, modified caution index and total score for the total set, and two subsets of items are reported in Table 3. The correlations among caution indices is .98 or higher for all three sets of items. The correlations of the caution indices with total test-score vary from being small negative (-.31) to small positive (.12). These correlations are consistent with other evidence that suggests that the sign and magnitude of the correlation between total scores and caution indices varies as a function of the skewness of the total test score distribution. No significant relationship was detected among the caution indices for differing sets of items. This was a very interesting finding never previously reported, which suggests either that the caution indices provide unreliable information or that the information is quite content specific. Results reported below are most consistent with the latter interpretation.

- - - - - - - - - - - - -

Insert Table 3 about here

- - - - - - - - - - - - -

The scatterplot for the modified caution index computed over all items with the total score is shown in Figure 1. The negative relationship between the modified caution index and total score can be seen in the scatterplot. The presence of high caution indices for individuals with low total scores could possibly result from a combination of factors including guessing correctly on one or more of the very difficult items, irregular study habits for low achievers, long-term absence from instruction, wide range of item difficulties on the test, teachers stressing problem solving activities which improves student performance on the type of mathematics exercises that are considered the most difficult for the nation, and teachers not covering parts of the curriculum that are in the NAEP assessment of mathematic achievement. A maximum value of .56 was given the individual with the most unusual response pattern. No one with a total correct score of 25 or more had a caution index higher than .40, whereas 35 of the people with total scores less than 25 had caution indices exceeding that value.

- - - - - - - - - - - - - - -

Insert Figure 1 about here

- - - - - - - - - - - - - - -

The scatterplot for the modified caution index computed for the arithmetic items with the total score is shown in Figure 2. One student performing very well on these items missed a very easy item contributing to the largest modified caution index of all at .81. The scatterplot reveals the negative relationship between the modified caution index value and the total score. Similar factors as reported earlier could possibly

help to explain the different response patterns.

- - - - - - - - - - - - - - -

Insert Figure 2 about here

- - - - - - - - - - - - - - -

An interesting change in the relationship can be seen in the scatterplot for the modified caution index computed for the geometry items with the total score for geometry items (Figure 3). A positive but near zero relationship is found between the modified caution index for geometry and the total geometry score. This change in relationship from negative to positive suggests the strength of the relationship between the modified caution index and the total test score is confounded by the nature of the total score distribution.

- - - - - - - - - - - - - - -

Insert Figure 3 about here

- - - - - - - - - - - - - - -

Of particular importance and seen in the scatterplot of modified caution indices computed for the arithmetic and geometry items given in Figure 4, is the small positive near zero relationship among the modified caution indices across the two domain of items. This result suggests that unusual response patterns for individuals are not consistent across item domains.

- - - - - - - - - - - - - - -

Insert Figure 4 about here

- - - - - - - - - - - - - - -

Since Sato's caution index was correlated .99 with the modified caution index, we will drop reporting of Sato's caution index in favor of the modified caution index which is bounded at 0 and 1. School variables

that were significantly correlated in the positive direction with the modified caution index for the total set of items were percent of community on welfare, percent of community farm workers, percent of community black, and number of students qualified to receive Title I and are reported in Table 4. School variables that were significantly correlated in the negative direction with the modified caution index for the total set of items were percent of community profession workers, percent of community skilled workers, and type of school.

- - - - - - - - - - - - - -

Insert Table 4 about here

- - - - - - - - - - - - - -

No significant gender differences were found with the modified caution index. However, the results all consistently show females as having lower caution values across the different item sets.

The Like Mathematics scale revealed no significant relationship with the caution values while the Usefulness of Mathematics scale showed a significant negative relationship with the caution index. Students perceiving mathematics as not very useful were students with high caution indices.

Since the caution indices were substantially correlated with the total score, partial correlations of school and student background were computed controlling for the respective total test score. The partial correlations are given in Table 5. Significant positive partial correlations were found with the modified caution index for the total set of items on the following variables: percent of community farm workers, percent of community black, and number of students qualified to receive Title I. Significant negative partial correlations were found with the modified caution index for the

total set of items on the following variables: percent of community skilled or unskilled workers, and the usefulness of Mathematics Scale.

- - - - - - - - - - - - -

Insert Table 5 about here

- - - - - - - - - - - - -

The results of the analyses of variance and of covariance (ANOVA and ANCOVA) are given in Table 6. For the total set of items and for the arithmetic items, the caution index varied significantly on the set of four variables investigated: community size, race-ethnicity, grade level, and census division. The ANCOVA results, where we controlled for the total test score, revealed only one variable with a significant main effect remaining. This significant effect was for the race-ethnicity variable indicating that the differences in caution indices were still significant between groups of students classified on this variable.

- - - - - - - - - - - - -

Insert Table 6 about here

- - - - - - - - - - - - -

The results from the ANOVA for the geometry set of items revealed that the caution index varied significantly on only the race-ethnicity variable. The ANCOVA results after controlling for the total geometry score revealed significant differences both on the race-ethnicity and the census division variable.

To illustrate the varying distributions of caution indices by total score for specific subgroups we prepared "Box and Whisker Plots." A box and whisker plot provides a graphical picture of the location of the 25^{th}, 50^{th}, and 75^{th} percentile score with the range of values plus the mean

value. The box represents the middle 50% of the values while the whisker or tails extending from the box represent the lower and upper quartile values. Values of 0 on the box and whisker plot indicate the probability occurrence as 1 chance out of 20 while an asterisk indicates the probability of occurrence at 1 chance out of 200.

The total test score interval for the total and arithmetic set of items were split into deciles while the test score interval for the geometry set of items was split into quintiles. Box and Whisker plots of the modified caution index for blacks and whites by their test performance level are given in Figures 5, 6, and 7 for the total arithmetic and geometry set of items, respectively. Blacks across each of three total test intervals for each decile have higher caution indices as can be seen in Figures 5, 6, and 7.

- - - - - - - - - - - - - - - - - - -

Insert Figures 5, 6, and 7 about here

- - - - - - - - - - - - - - - - - - -

Box and whisker plots of the modified caution index values across the test score performance levels for males and females are given in Figures 8, 9, and 10 for the total, arithmetic and geometry set of items, respectively. An interesting pattern of results can be seen in each of these figures; that is, the females of a low ability level have lower caution values than males of equal ability while females of a high ability level have greater caution indices than males of equal ability. The midspread (i.e., the distance from the 25th to the 75th percentile) of caution indices for the students in the first decile are about one and a half times the midspread throughout the remaining deciles.

- - - - - - - - - - - - - - - - - - -

Insert Figures 8, 9, and 10 about here

- - - - - - - - - - - - - - - - - - -

Box and whisker plots of the modified caution index values across the test score performance levels for two groups formed based on a median split of the Usefulness of Mathematics scale are given in Figures 11, 12, and 13 for the total, arithmetic and geometry set of items, respectively. At the extremes of the test score intervals higher caution values can be seen for students expressing little usefulness of mathematics compared to students who expressed a great amount of usefulness of mathematics. No apparent differences between groups formed based on the usefulness of mathematics could be seen in the middle region of the test score interval.

- -

Insert Figures 11, 12, and 13 about here.

- -

Box and whisker plots of modified caution index values were also obtained and inspected for a high and a low group of students based on a median split of the Like Mathematics scale. No major differences were found that would necessitate their figures to be included.

Since the differences in modified caution index values noted on the box and whisker plots for blacks and whites could be influenced by total test score, we evaluated the residuals remaining after regressing the modified caution index on total test score. Box and whisker plots of the residuals across the test score intervals are given for blacks and whites in Figures 14, 15, and 16 for the total, arithmetic and geometry set of items. The pattern of blacks higher residuals can be seen in the figures across all three test score intervals. This pattern of differences in response patterns between blacks and whites after controlling for test score indicate that groups of students differ in their response patterns even after adjusting for differences in total correct scores. On average,

the response patterns for black students are less consistent with the national norm than are the response patterns of white students with comparable total number correct scores.

- -

Insert Figures 14, 15, and 16 about here

- -

School and Census Division Differences on Achievement and on the Modified Caution Index

To determine the variation in achievement and modified caution indices that can be explained by census division and school differences within census divisions we divided the variation among the pupil's achievement test scores and modified caution index values into a between-census division and a within-census division component. Table 7 shows the percent of variation explained for pupil's mathematics achievement and modified caution index values on the total, arithmetic and geometry set of items.

- - - - - - - - - - - - - -

Insert Table 7 about here

- - - - - - - - - - - - - -

We will now focus on the results of the variance decomposition of the mathematics achievement total correct scores and the modified caution index for the total, arithmetic and geometry set of items. Between census division variance on achievement ranged from 2.9% of total variance on the arithmetic items to 3.5% on arithmetic items. Schools within-census division achievement variance ranged from a high of 25.5% on geometry items to a low of 22.5% on arithmetic items. For the total set of items 4% of the total variance was explained by between census division variance and 28.2% explained by schools within census division. Clearly, there are sizeable

between school differences in achievement, a result that could have been predicted with great confidence without the analysis. But it is also clear that the within school variability in achievement accounts for a much larger fraction of the total variation.

Between census division variance on modified caution indices ranged from .05% of the total variance on arithmetic items to .43% on geometry items. Schools within census division modified caution index variance ranged from a high of 7.8% on arithmetic items to a low of 1.9% on geometry items. Schools within census division explained 11.8% of the total variance while only .8% was explained by between census divisions. The between school variability in the caution indices is noticeably less than that for total scores. Nonetheless, there are sizeable between school differences in caution indices for the total set of items and for the set of arithmetic items. These between school differences suggest significant variation between schools in content emphasis.

The modified caution indices for all three tests were used as dependent variables in a hierarchical ANOVA. The first factor is the nine different census divisions of the nation. School is the second factor. This factor is nested in census divisions, while students, the third factor, are nested within schools within census divisions. The results of these analyses are summarized in Table 8. The schools within census divisions 3, 6, 7, and 9 have significantly different modified caution indices on the total item set.

- - - - - - - - - - - - - -

Insert Table 8 about here

- - - - - - - - - - - - - -

The mean caution indices on all tests for the 13 schools in census division 6 and 15 schools in census division 7 are reported in Tables 9 and 10 respectively. The range of the school mean modified caution indices is .15 to .27 on all items. The relatively wide range of modified caution indices for schools within these census divisions suggests a high degree of variability of item response patterns. These large differences as noted in Tables 9 and 10 may well be a function of the curriculum. The school effects are significant for all indices, which reveals that certain schools may not have covered segments of the content sample on the test, or that they may have given less than typical emphasis to some of the content.

- - - - - - - - - - - - - - - - - -

Insert Tables 9 and 10 about here

- - - - - - - - - - - - - - - - - -

The significant schools-within-census division effects denote the high degree of variability of student performance in schools within certain census divisions of the country. Curriculum offerings may very well contribute to these large differences. To explore this possibility, we conducted a more detailed analysis of response patterns of students at schools within census division 6 and 7 following the approach used by Harnisch and Linn (1981). This study was designed to identify the subset of items which contribute most to the caution indices for the use to describe unique patterns of performance by item content. Various patterns of performance suggest differences in content coverage that make the test less appropriate for some schools than others.

Schools from census division 6 and 7 with 10 or more students were evaluated for unusual response patterns. The p-values, i.e., the proportion of students who answered an item correctly, were computed for

each school on each of the 68 items. Since school mean performance on the test is directly related to the p-values on the items, a linear regression was performed on the p-values for each school with the p-values from the nation. The regression equation was used to compute the expected proportion correct on each item for each school. Residual scores were computed simply by subtraction of expected from observed proportion correct on each item for each school.

Items were categorized in terms of their content and format in order to find clues about the possible reasons for the large differences in the residuals. The mean of the residuals for each category was then standardized by dividing by the standard error of estimate. Finally, the standardized mean residuals were multiplied by the square root of the number of items in the content category as a means of weighting the standardized mean residuals according to the number of items in the category. The resulting weighted standardized mean residuals, which are analogous to critical ratios, were used to compare the items in different categories. These results are reported in Tables 11 and 12 for the two census divisions.

- - - - - - - - - - - - - - - - - - -

Insert Tables 11 and 12 about here

- - - - - - - - - - - - - - - - - - -

An entry in Table 11 or 12 greater than 2.0 in absolute value indicates that items in that particular category are much easier or much harder for students in that school than would be expected from their overall performance and the relative difficulty of these items for the national sample as a whole. Five of the content categories for schools within census division 6 have entries in Table 11 greater than the 2.0

absolute value for one or more schools. Seven of the content categories for school within census division 7 have entries in Table 12 greater than the 2.0 absolute value for one or more schools. The one category of items in which two or more schools from either census division 6 or 7 had large entries was sign numbers. For schools within census division 6, the large positive entry (3.88) for School 2 for the sign numbers stands in marked contrast to the large negative values for Schools 3 and 4 in this category (-2.45 and -2.23 respectively). This suggests the hypothesis that the use of sign numbers may be quite common in School 2 but rare in Schools 3 and 4. Similar hypotheses are suggested by the other large values in Table 11 and Table 12.

Booklet 5

A stem-and-leaf display of the percentage correct values on the 63 cognitive items used in Booklet 5 is given in Table 13. The item difficulties ranged in percentage correct from 2 to 95%. Twenty of the items are answered correctly by greater than 70% of the test takers while nineteen of the items are answered correctly by fewer than 40% of the test takers. Appendix C provides for each item the percent correct, the point-biserial, Sato's caution index, and the modified caution index. Two items were identified with modified caution indices over .5 suggesting that an unexpected large number of low performing students are answering items correctly. The means, standard deviations, and correlations among the four-item statistics are given in Appendix D. The caution indices correlated as low or lower with the percentage correct as compared with the correlation of the point-biserial with the percentage correct. The correlation of the caution indices with the point-biserial was a high negative value. The extremely high correlations among the caution indices

for the items (.00) is not surprising as was noted earlier with reference to Book 4 items. It can be stated that the two indices yield nearly identical information about the items.

- - - - - - - - - - - - - - -

Insert Table 13 about here

- - - - - - - - - - - - - - -

The intercorrelations among Sato's caution index, modified caution index and total score for the total set and two subsets of items are given in Table 14. The correlation among caution indices is .99 for all three sets of items. The correlations of the caution indices with total test score varies from being small negative (-.21) to small positive (.01). The correlations between the caution indices and total test score for arithmetic and geometry items are nearly zero. Since the relationship of caution indices with test score overall items is negative, this suggests that the non-arithmetic and non-geometry items are contributing substantially to this correlation. The relationship of caution indices across different item subsets was nearly zero. This concurs with our finding from Book 4 and suggests that the information is quite content specific. The results reported for Book 4 and the below are most consistent with this interpretation.

- - - - - - - - - - - - - - -

Insert Table 14 about here

- - - - - - - - - - - - - - -

The scatterplot for the modified caution index computed over all items with the total test score is shown in Figure 17. The small negative relationship of modified caution values with the total test score can be seen in the scatterplot. As noted earlier, and worth repeating, is the

notion that high caution indices for individuals can result from a combination of many different factors. These large caution indices possibly result from a student guessing correctly on one or more of the very difficult items, irregular study habits for low achievers, long-term absence from instruction, wide range of item difficulties on the test, teachers stressing problem solving activities which improve student performance on the type of mathematics exercises that are considered the most difficult for the nation, and teachers not covering parts of the curriculum that are in the NAEP assessment of mathematics achievement. A maximum value of .58 was given the individual with the most unusual response pattern. No one with a total correct score of 28 or more had a caution index higher than .40, whereas 6 of the people with total scores less than 28 had caution indices exceeding that value.

- - - - - - - - - - - - - - -

Insert Figure 17 about here

- - - - - - - - - - - - - - -

The scatterplot for the modified catuion index computed for the arithmetic items with the total score is shown in Figure 18. One student answering only two items correctly and perhaps nearly the most difficult arithmetic items had the largest modified caution index value of .88. The scatterplot reveals a broad range of modified caution values across the test score interval with only a few unusually high values spread across the test score distribution. A host of factors as noted earlier could possibly help to explain the different response patterns.

- - - - - - - - - - - - - - -

Insert Figure 18 about here

- - - - - - - - - - - - - - -

The scatterplot for the modified caution index computed for the geometry items with the total score for geometry items is given in Figure 19. The scatterplot reveals the near zero relationship between the modified caution index and the total geometry test score. Modified caution index values exceeding .90 are found for five students scoring nine or above on the geometry items. The near zero relationship as seen from the scatterplot indicates that the caution index values are providing us with new information not contained in the total test score. An interesting area of research is trying to understand the factors associated with students that have large caution values or schools with large numbers of students who have large caution values.

- - - - - - - - - - - - - - -

Insert Figure 19 about here

- - - - - - - - - - - - - - -

No systematic relationship was found between the modified caution index computed on arithmetic items and the modified caution index computed on geometry items. The scatterplot that reveals this near zero relationship is given in Figure 20. An individual with a large caution value on one subset of items is not necessarily going to have a large caution value on a different set of items.

- - - - - - - - - - - - - - -

Insert Figure 20 about here

- - - - - - - - - - - - - - -

Since the relationship between Sato's caution index and the modified caution index was nearly one, we will continue reporting only the modified caution index since it is bounded at 0 and 1. The correlations between a set of school, student background and affective domain variables and the

modified caution index and total test score are given in Table 15 for the total, arithmetic and geometry sets of items. Sex is found to be the variable most associated with the caution indices. Females displayed consistently lower caution values across the different item sets. The two school variables found to be associated with modified caution index values are percent of community on welfare and percent of community black, each in a positive manner.

- - - - - - - - - - - - - - -

Insert Table 15 about here

- - - - - - - - - - - - - - -

For all practical purposes, no significant relationship was found between the affective domain scales and the modified caution index. However, the validity coefficients for all the affective domain variables with the respective total test score are all positive and range from .06 to .31. The easiness of mathematics scale consistently correlates .20 or higher with each of the total test scores.

The results of the ANOVA and ANCOVA are reported in Table 16. Significant differences in caution indices for the total set of items were found for groups formed based on their race-ethnicity and grade level variables. No significant differences were found between groups for the modified caution index computed on the arithmetic items. On the geometry items, the three variables found to indicate large between-group differences on the modified caution index were community size, race-ethnicity and census division. The ANCOVA results on the geometry items, where we controlled for the total geometry test score, revealed significant between-group differences on each of the three variables noted above. On the

other hand, the ANCOVA results on the total set of items revealed no significant between group differences after controlling for the total test score.

- - - - - - - - - - - - - - -

Insert Table 16 about here

- - - - - - - - - - - - - - -

The distributions of caution indices by total test score for specific subgroups were obtained and inspected. Box and whisker plots were prepared to illustrate the conditional distributions of caution indices for the total test score interval split into deciles for the arithmetic and total set of items while the geometry test score interval was split into quintiles. Only two of these plots will be included.

Box and whisker plots of the modified caution index values across the test score performance levels for males and females are given in Figures 21 and 22 for total and arithmetic set of items, respectively. The pattern of results shown in Figure 21 is consistent for the males across the test performance levels with males having greater caution indices. For the arithmetic set of items, a similar pattern of results are found at the extreme performance levels while some moderation of differences in caution indices occurs at the middle range of the performance levels.

- - - - - - - - - - - - - - - - - - -

Insert Figures 21 and 22 about here

- - - - - - - - - - - - - - - - - - -

School and Census Division Differences on Achievement and on the Modified Caution Index

To determine the variation in achievement and modified caution indices that can be explained by census division and school differences within

census divisions we divided the variation among the pupil's achievement test scores and modified caution index values into a between-census division and a within-census division component. Table 17 shows the percent of variation explained for pupil's mathematics achievement and modified caution index values on the total, arithmetic and geometry set of items.

- - - - - - - - - - - - - - -

Insert Table 17 about here

- - - - - - - - - - - - - - -

The variance decomposition of the mathematics achievement total correct score for the total, arithmetic and geometry set of items into a between census divisions component and a schools within census divisions revealed similar percentage patterns as were found with Book 4. Between census division variance on achievement ranged from 2.0% on the geometry items to 4.3% on the arithmetic items. Over two times as much of the variance in achievement was explained for the arithmetic items compared to the geometry items with the schools within census divisions component of variation. For the total set of items 5.2% of the total variance was explained by between census division variance and 23.3% explained by schools within census division. In each of the subsets and also for the total set of items the percentage of variance occurring within schools is much greater than between schools, revealing a wide range of individual differences within schools.

The schools within census division modified caution index variance was nearly the same on the arithmetic and geometry set of items, 2.7% and 2.3%, respectively. For the total set of items, schools within census division explained 3.5% of the total variance. Differences at the between school level on the caution indices suggest significant variation between schools in emphasis given to the subject matter.

The modified caution indices for all three tests were used as dependent variables in a hierarchical ANOVA. The first factor is the nine different census divisions of the nation. School is the second factor. This factor is nested in census divisions, while students, the third factor, are nested within schools within census divisions. The results of these analyses are summarized in Table 18. The schools within census divisions 2 and 9 have significantly different modified caution indices on the total item set. For the geometry item set, schools within census divisions 1 and 3 have significantly different modified caution indices.

- - - - - - - - - - - - - - -

Insert Table 18 about here

- - - - - - - - - - - - - - -

The mean caution indices on all tests for the 8 schools in census division 1 and 21 schools in census division 9 are reported in Table 19 and 20, respectively. The range of the school mean modified caution indices is .03 to .25 on the geometry set of items for schools in census division 1 and .11 to .28 on the arithmetic set of items for schools within census division 9. The relatively wide range of modified caution indices for schools within these census divisions suggests a high degree of variability of item response patterns. These large differences as noted in Tables 19 and 20 may well have been a function of the curriculum. The school effects are significant for all indices, which reveals that certain schools may not have covered segments of the content sampled on the test, or that they have given less than typical emphasis to some of the content.

- - - - - - - - - - - - - - - - - - -

Insert Tables 19 and 20 about here

- - - - - - - - - - - - - - - - - - -

The significant schools-within-census division effects denote the high degree of variability of student performance in schools within certain census divisions of the country. Curriculum offerings may very well contribute to these large differences. To explore this possibility, we conducted a more detailed analysis of response patterns of students at schools within census division 1 and 9 following the approach used by Harnisch and Linn (1981). This analysis was designed to identify subsets of items which contribute most to the caution indices for the respective schools. The analyses of these subsets of items were used to describe unique patterns of performance by item content. Various patterns of performance suggest differences in content coverage that make the test less appropriate for some schools than others.

Schools from census divisions 1 and 9 with 10 or more students were evaluated for unusual response patterns. The p-values, i.e., the proportion of students who answered an item correctly, were computed for each school on each of the 68 items. Since school mean performance on the test is directly related to the p-values on the items, a linear regression was performed on the p-values for each school with the p-values from the nation. The regression equation was used to compute the expected proportion correct on each item for each school. Residual scores were computed simply by subtraction of expected from observed proportion correct on each item for each school.

Items were categorized in terms of their content and format in order to find clues about the possible reasons for the large differences in the residuals. The means of the residuals for each category were then standardized by dividing by the standard error of estimate. Finally, the standardized mean residuals were multiplied by the square root of

the number of items in the content category as a means of weighting the standardized mean residuals according to the number of items in the category. The resulting weighted standardized mean residuals, which are analogous to critical ratios, were used to compare the items in different categories. These results are reported in Tables 21 and 22 for census divisions 1 and 9, respectively.

- - - - - - - - - - - - - - - - - - - -

Insert Tables 21 and 22 about here

- - - - - - - - - - - - - - - - - - - -

An entry in Table 21 or 22 greater than 2.0 in absolute value indicates that items in that particular category are much easier or much harder for students in that school than would be expected from their overall performance and the relative difficulty of these items for the national sample as a whole. Five of the content categories for schools within census division 1 have entries in Table 21 greater than the 2.0 absolute value for one or more schools. Story problems is the only content category that does not have entries greater than 2.0 absolute value for one or more schools within census division 9. The fractions-to-decimals category of items had three or more schools from either census division 1 or 9 with large entries. For schools within census division 1, the large positive entry (2.81) for School 1 for fractions-to-decimals stands in marked contrast to the large negative value for Schools 5 and 8 in this category (-3.09 and -3.35, respectively). This suggests that the hypothesis that the practice of transforming fractions to decimals may be quite common in School 1 but rare in Schools 5 and 8. Similar hypotheses are suggested by the other large values in Tables 21 and 22. Thus, one would expect that special emphasis is placed on performing calculations in Schools 14, 16 and 19 from census division 9.

SUMMARY AND IMPLICATIONS

Sato's caution index and the modified caution index were each computed for the students and the test items from two test booklets from the 1978 math assessment for the 13 year old sample. The correlation between the modified caution index and Sato's caution index was nearly one both for the indices computed for the students and the problems. The remainder of the analyses focused on reporting the variables associated with the modified caution index. Comparisons of subgroup caution index values were used to identify student background characteristics that were associated with unusual response patterns.

Near zero correlations were found between the caution indices computed for different subsets of items. Further analyses revealed that the information contained in the modified caution index was quite content specific rather than providing unreliable information about an individual.

Significant gender differences were found on the caution indices revealing that females tend to have smaller caution values across the total test score interval. Significant race-ethnicity differences were found on the caution indices with blacks having larger caution values across the total test score interval.

Significant differences in caution indices, namely lower values, were found associated with students having a perception that mathematics was not very useful. This relationship was quite apparent throughout the total test score interval.

Differences in response patterns at the school level were identified and specific subsets of items were identified as contributing to these large caution values. The association of subsets of items with large caution indices at particular schools has potential diagnostic value.

While numerous reasons may be given to explain the school differences in caution indices it is quite plausible that the differences in response patterns result from variability in content coverage and emphasis. This possibility deserves further study.

From a research point of view, it would be most interesting and informative to extend the analysis to the National Assessment of Educational Progress sample of 9- and 17-year-olds to make further comparisons of subgroups on the caution index values as well as identifying student background characteristics that are associated with students having unusual response patterns.

Differences in response patterns found in this study indicate that there are individuals for whom the total test score conceals systematic deviations from the typical pattern of responses and may be misleading. The results also suggest that there are schools with curricula that do not match the test content. Formation of a contingency table as suggested by Sato (1975) for the caution index along with the total test score can be used to identify students who need more study, who make careless mistakes, who possess sporadic study habits or insufficient readiness or who are doing everything fine. Schools can similarly be identified as having in general one or more of the above categories of students.

Notes

1. Tatsuoka, K., & Tatsuoka, M. M. Detection of aberrant response patterns and their effects on dimensionality. (Research Report 80-4) Urbana, IL: University of Illinois, Computer-based Education Research Laboratory, 1980.

2. Public Use Data Tapes can be purchased from:

   Department of User Services
   National Assessment of Educational Progress
   Education Commission of the States
   1860 Lincoln Street, Suite 700
   Denver, Colorado 80295

3. Tatsuoka, M. M. Recent psychometric developments in Japan: Engineers grapple with educational measurement problems. Paper presented at the ONR Contractors Meeting on Individualized Measurement, Columbia, MO, 1978.

References

Alderman, D. L., Swinton, S. S., & Braswell, J. S. Assessing basic arithmetic skills and understanding across curricula: Computer-assisted instruction and compensatory education. The Journal of Children's Mathematical Behavior, 1979, 2, 3-28.

Donlon, T. F., & Fischer, F. E. An index of an individual's agreement with group-determined item difficulties. Educational Psychological Measurement, 1968, 28, 105-113.

Drasgow, F. Statistical indices of the appropriateness of aptitude test scores. Unpublished doctoral dissertation, University of Illinois, Urbana-Champaign, 1978.

Finn, J. D. Multivariance: Univariate and multivariate analysis of variance, covariance, regression and repeated measures. Chicago: International Educational Services, 1977.

Guttman, L. The quantification of a class of attributes: A theory and method of scale construction. In P. Horst, P. Wallin, & L. Guttman (Eds.), The prediction of personal adjustment. New York: Social Science Research Council, Committee on Social Adjustment, 1941.

Harnisch, D. L., & Linn, R. L. Analysis of item response patterns: Questionable test data and dissimilar curriculum practices. Journal of Educational Measurement, 1981, 18, 133-146.

House, E. R., Glass, G. V., McLean, L. D., & Walker, D. F. No simple answer: Critique of the follow-through evaluation. Harvard Educational Review, 1978, 48, 128-160.

Kane, M. T., & Brennan, R. L. Agreement coefficients as indices of dependability for domain-referenced tests. Applied Psychological Measurement, 1980, 4, 105-126.

Levine, M. V., & Rubin, D. B. Measuring the appropriateness of multiple choice test scores. Journal of Educational Statistics, 1979, 4, 269-290.

Madaus, G. F., Kelleghan, T., Rakow, E. G., & King, D. J. The sensitivity of measures of school effectiveness. Harvard Educational Review, 1979, 49, 207-230.

Moore, R. P., Chromy, J. R., & Rogers, W. T. The national assessment approach to sampling. Denver, Colo.: Education Commission of the States, National Assessment of Educational Progress, 1974.

National Center for Education Statistics. Three national assessments of science: Changes in achievement, 1969-77. Science Report No. 08-S-00, Denver, Colo.: National Assessment of Educational Progress, 1978.

Nie, N. H., Hull, C. H., Jenkins, J. G., Steinbrenner, K., & Bent, D. SPSS: Statistical package for the social sciences. St. Louis: McGraw Hill Book Company, 1975.

Porter, A. C., Schmidt, W. H., Floden, R. E., & Freeman, D. J. Impact on what?: The importance of the content covered. (Research Series Report No. 2). East Lansing: Michigan State University, The Institute for Research on Teaching, 1978.

SAS User's Guide, 1979 Edition. Raleigh, North Carolina: SAS Institute Inc., 1979.

Sato, T. [The construction and interpretation of S-P tables.] Tokyo: Meiji Tosho, 1975.

Sato, T. Similarity of some indices of item response patterns (Research Memorandum E181-1). Tokyo: Computer and Communications System Research Laboratories, Nippon Electric Co., Ltd., 1981.

Van der Flier, H. Environmental Factors and deviant response patterns. In Y. H. Poortinga (Ed.), Basic problems in cross-cultural psychology. Amsterdam: Swets and Seitlinger, B. V., 1977.

Wright, B. D. Solving measurement problems with the Rasch model. Journal of Educational Measurement, 1977, 14, 97-116.

Wright, D. Introduction to the National Assessment of Educational Progress Public Use Data Tapes. Denver, Colo.: Education Commission of the States, National Assessment of Educational Progress, 1981.

Exhibits 1 and 2

Exhibit 1

Book 4 Affective Domain Items

Like Scale Items

1. Feel like taking mathematics tests
2. Feel like doing mathematics homework
3. Feel like helping a classmate do mathematics
4. Feel like playing mathematics games
5. Feel like listening to teacher explain math
6. Feel like watching teacher work at the board
7. Feel like using hand calculators in math
8. Feel like using a computer in math

Usefulness Scale Items

1. Feel useful to take mathematics tests
2. Feel useful doing mathematics homework
3. Feel useful halping a classmate do math
4. Feel useful to play mathematics games
5. Feel useful to listen to teacher explain math
6. Feel useful to watch teacher work at the board
7. Feel useful to use hand calculators in math
8. Feel useful to use a computer in math

Exhibit 2

Book 5 Affective Domain Items

Students rated each of the mathematics activities below with reference to like for, importance, and easiness.

1. Solving word problems
2. Working with fractions
3. Estimating answers to problems
4. Measuring lengths, weights or volume
5. Working with metric measures
6. Doing proofs

Tables 1 - 22

Table 1

S-P Table for 18 Examinees and 5 Items

(Hypothetical Example)

Examinee i	Item j 1	2	3	4	5	Examinee Total $n_{i\cdot}$	Sato's Caution Index C_i	Modified Caution Index C_i^*
1	1	1	1	1	0	4	.00	.00
2	1	1	1	0	1	4	.65	.33
3	1	1	1	0	0	3	.00	.00
4	1	1	0	1	0	3	.16	.08
5	1	1	0	0	1	3	.65	.31
6	1	0	1	0	1	3	1.13	.54
7	1	1	0	0	0	2	.00	.00
8	1	1	0	0	0	2	.00	.00
9	1	0	1	0	0	2	.44	.23
10	1	0	0	1	0	2	.59	.31
11	0	1	1	0	0	2	.74	.39
12	0	1	0	1	0	2	.88	.47
13	1	0	0	0	0	1	.00	.00
14	1	0	0	0	0	1	.00	.00
15	0	1	0	0	0	1	.45	.22
16	0	0	1	0	0	1	1.14	.56
17	0	0	0	1	0	1	1.36	.67
18	0	0	0	1	0	1	1.36	.67
Item Total $n_{\cdot j}$	12	10	7	6	3			
Sato's Caution Index C_j	.13	.28	.42	.95	.21			
Modified Caution Index C_j^*	.14	.14	.21	.50	.13			

Table 2

Stem and Leaf Display of p Values for Cognitive Items--
Book 4, 13-year olds, 1978
(N = 68)

Stem	Leaf								Frequency
90	7.17								1
90	4.54	1.10							2
80	8.14	6.83	6.62	5.02					4
80	4.16	4.08	3.34	2.81					4
70	9.85	9.48	9.44						3
70	1.69	1.48							2
60	8.24	5.90	5.78	5.12					4
60	4.87	3.97							2
50	7.12	7.00	5.31						3
50	4.78	4.37	3.10	0.96	0.55	0.14			6
40	9.65	7.97	6.37	5.84					4
40	4.36	3.33							2
30	9.68	8.65	5.70	5.13					4
30	4.39	3.28	3.03	2.87	2.50	1.68	1.23	0.57	8
20	9.13	8.48	8.44	6.14	5.77	5.24			6
20	4.33	4.29	3.43	3.14	2.65				5
10	9.45	9.45	7.28	6.78	6.50				5
10									0
0	7.35								1
0	3.90	2.05							2
									68

Table 3

Means, Standard Deviations, Skewness, and the Pearson Product Moment Correlations Among Sato's Caution Index, Modified Caution Index and the Total Test Score for the Total Set and Two Subsets of Items

Book 4, 13 year olds, 1978

(N = 1219)

Variable	Variable 1	2	3	4	5	6	7	8	9
All Items									
1. Sato's Caution Index	1.00	.99	-.19	.74	.74	-.15	.51	.51	-.01
2. Modified Caution Index		1.00	-.31	.74	.76	-.27	.48	.49	-.11
3. Total Test Score			1.00	-.17	-.31	.95	.05	-.02	.81
Arithmetic Items									
4. Sato's Caution Index				1.00	.98	-.17	.07	.08	-.08
5. Modified Caution Indes					1.00	-.32	.07	.08	-.18
6. Total Test Score						1.00	.03	-.02	.64
Geometry Items									
7. Sato's Caution Index							1.00	.99	.12
8. Modified Caution Index								1.00	.04
9. Total Test Score									1.00
Mean	.35	.18	33.00	.29	.15	16.79	.40	.20	8.30
Standard Deviation	.13	.07	10.56	.16	.09	6.20	.23	.12	3.28
Skewness	.91	1.10	.28	1.16	1.26	.31	.70	.74	.58

Table 4

Pearson Product Moment Correlations* of School, Student Background, and Affective Domain Variables with Modified Caution Index and the Total Test Score for the Total Set and Two Subsets of Items

Book 4, 13 year olds, 1978

(N = 1219)

Variable	All Items		Arithmetic Items		Geometry Items	
	Modified Caution	Total Score	Modified Caution	Total Score	Modified Caution	Total Score
1. Percent of Community unskilled	-.04	-.05	-.02	-.05	-.02	-.07
2. Percent of community not employed	.07	-.14	.05	-.12	.01	-.14
3. Percent of community on welfare	.17	-.32	.16	-.29	.03	-.27
4. Percent of community farm workers	.10	-.10	.06	-.10	.04	-.03
5. Percent of community black	.19	-.28	.17	-.26	.05	-.21
6. Percent of community professional workers	-.10	.29	-.10	.25	.01	.27
7. Percent of community skilled workers	-.12	.18	-.09	.19	-.06	.11
8. Number of students qualified to receive Title I	.13	-.24	.12	-.23	-.01	-.16
9. School enrollment size	.01	-.01	.01	-.02	-.02	.01
10. Sex[a]	-.05	-.06	-.05	-.01	-.02	-.09
11. Family owns calculator[b]	-.09	.19	-.05	.17	-.02	.13
12. Type of school[c]	-.09	.14	-.05	.16	-.06	.06
13. Like mathematics scale	.03	-.04	.03	-.03	.03	-.01
14. Usefulness of mathematics scale	-.14	.24	-.08	.24	-.06	.15

[a]Sex is coded 1 for males and 2 for females.

[b]Family owns calculator is coded 0 for no and 1 for yes.

[c]Type of school is coded 1 for public and 2 for private Catholic.

*Correlations greater in absolute magnitude than .07 are significant at $p < .01$.

Table 5

Partial Correlations* of School, Student Background, and Affective Domain Variables with Modified Caution Index Controlling for the Total Test Score for the Total Set and Two Subsets of Items

Book 4, 13 year olds, 1978
(N = 1219)

Variable	All Items	Arithmetic Items	Geometry Items
1. Percent of Community unskilled	-.07	-.05	-.02
2. Percent of community not employed	.03	.01	.02
3. Percent of community on welfare	.06	.06	.04
4. Percent of community farm workers	.09	.05	.05
5. Percent of community black	.11	.08	.06
6. Percent of community professional workers	-.01	-.02	-.01
7. Percent of community skilled workers	-.07	-.03	-.07
8. Number of students qualified to receive Title 1	.07	.05	.00
9. School enrollment size	.01	.00	-.02
10. Sex[a]	-.04	-.04	-.01
11. Family owns calculator[b]	-.04	.00	-.03
12. Type of school[c]	-.06	.00	-.07
13. Like mathematics scale	.02	.01	.03
14. Usefulness of mathematics scale	-.08	-.01	-.07

[a]Sex is coded 1 for males and 2 for females.

[b]Family owns calculator is coded 0 for no and 1 for yes.

[c]Type of school is coded 1 for public and 2 for private Catholic.

*Correlations greater in absolute magnitude than .07 are significant at $p < .01$.

Table 6

Analysis of Variance Summary Results for the Modified Caution Index with the Test Score as a Covariate for the Total Set and Two Subsets of Items

Book 4, 13 year olds, 1978
(N = 1219)

		All Items		Arithmetic Items		Geometry Items	
Source	df	Mean Square	F	Mean Square	F	Mean Square	F
ANOVA							
Main Effect							
Community size	3	.002	4.58**	.021	2.84*	.034	2.52
Residual	1214	.005		.007		.014	
ANCOVA							
Covariate	1	.573	129.96**	.906	135.72**	.031	2.29
Main Effect							
Community Size	3	.010	2.18	.004	.58	.033	2.44
Residual	1213	.004		.007		.013	
ANOVA							
Main Effect							
Race-Ethnicity	3	.123	26.73**	.124	17.37**	.056	4.18**
Residual	1214	.005		.007		.014	
ANCOVA							
Covariate	1	.573	132.76**	.906	137.68**	.031	2.31
Main Effect							
Race-Ethnicity	3	.047	10.91**	.042	6.43**	.074	5.50**
Residual	1213	.017		.007		.013	
ANOVA							
Main Effect							
Grade Level	4	.046	9.58**	.028	3.79**	.016	1.15
Residual	1213	.005		.007		.014	
ANCOVA							
Covariate	1	.573	129.90	.906	135.88**	.031	2.29
Main Effect							
Grade Level	4	.008	1.74	.007	1.06	.023	1.68
Residual	1212	.004				.014	
ANOVA							
Main Effect							
Census Division	8	.015	3.19**	.014	1.86	.023	1.69
Residual	1209	.005		.007		.014	
ANCOVA							
Covariate	1	.573	130.23**	.906	136.31**	.031	2.30
Main Effect							
Census Division	8	.008	1.76	.010	1.50	.027	1.96**
Residual	1208	.004		.007		.013	

*p < .05
**p < .01

Table 7

Percentages of Variation for the Components in

Mathematics Achievement and on the Modified Caution Index

Book 4, 15 year olds, 1978

	Between Census Division	Schools Within-Census Division	Error
Mathematics Achievement			
All items	3.97	28.22	67.81
Arithmetic items	2.92	25.51	71.57
Geometry items	3.46	22.51	74.57
Modified Caution Index			
All items	.76	11.77	87.47
Arithmetic items	.03	7.83	92.14
Geometry items	.43	1.89	97.68

Table 8

Summary of F Ratios for Hierarchical Analyses of Variance for the Modified Caution Index on the Total Set and Two Subsets of Items

Book 4, 13 year olds, 1978

Effect	All Items	Arithmetic Items	Geometry Items
Census Division	5.58*	2.65*	2.47
Schools within Census Division	3.13*	2.22*	1.36*
Schools within Census Division 1	2.27	3.06*	1.31
Schools within Census Division 2	1.57	1.43	1.17
Schools within Census Division 3	3.69**	2.88**	1.07
Schools within Census Division 4	1.13	1.31	1.36
Schools within Census Division 5	1.58	1.19	.82
Schools within Census Division 6	4.42**	3.41**	1.83
Schools within Census Division 7	5.40**	4.42**	1.70
Schools within Census Division 8	1.36	.68	.88
Schools within Census Division 9	5.11**	1.66	2.03*

*$p < .01$

**$p < .001$

Table 9

Means and Standard Deviations on Modified Caution Indices on the Total Set and Two Subsets of Items for Schools Within Census Division 6

Book 4, 13 year olds, 1978

School Number	Sample Size	All Items		Arithmetic Items		Geometry Items	
		Mean	S.D.	Mean	S.D.	Mean	S.D.
1	17	.17	.06	.15	.09	.18	.10
2	15	.19	.05	.18	.07	.21	.10
3	15	.17	.08	.17	.10	.18	.09
4	13	.19	.07	.17	.10	.17	.13
5	14	.18	.08	.14	.11	.24	.11
6	14	.23	.07	.19	.09	.30	.15
7	14	.17	.06	.15	.08	.18	.10
8	13	.22	.09	.20	.14	.26	.13
9	18	.15	.05	.12	.07	.19	.11
10	16	.16	.07	.12	.06	.23	.17
11	19	.18	.06	.15	.09	.18	.10
12	15	.27	.13	.26	.14	.26	.12
13	13	.25	.13	.22	.13	.23	.12

Table 10

Means and Standard Deviations on Modified Caution Indices on the Total Set and Two Subsets of Items for Schools Within Census Division 7

Book 4, 13 year olds, 1978

School Number	Sample Size	All Items		Arithmetic Items		Geometry Items	
		Mean	S.D.	Mean	S.D.	Mean	S.D.
1	16	.20	.09	.16	.08	.26	.11
2	16	.16	.07	.11	.08	.15	.08
3	15	.19	.06	.16	.07	.24	.15
4	15	.19	.09	.14	.09	.27	.15
5	14	.16	.08	.10	.04	.27	.17
6	20	.15	.05	.12	.07	.20	.12
7	13	.15	.06	.14	.11	.19	.10
8	10	.17	.07	.16	.07	.17	.12
9	15	.16	.08	.11	.05	.24	.16
10	15	.19	.08	.14	.08	.24	.16
11	15	.23	.11	.18	.08	.25	.16
12	14	.25	.11	.22	.12	.27	.11
13	16	.16	.07	.15	.07	.18	.09
14	14	.27	.09	.23	.13	.24	.13
15	25	.25	.08	.25	.10	.23	.17

Table 11

Weighted Standardized Mean Residuals of Within School Item Difficulties by Content Category for Schools Within Census Division 6

Book 4, 13 year olds, 1978

Content Category	School 1	2	3	4	5	6	7	8	9	10	11	12	13
Story Problems	.50	- .99	.35	.02	- .72	.29	.71	- .58	- .21	- .30	.71	1.13	1.03
Calculation	- .33	.27	-1.02	- .77	-1.58	.73	.24	1.10	.48	.78	-1.21	.11	1.10
Geometry-- Definition	-1.52	- .40	1.17	1.74	1.15	.16	- .52	.77	1.90	1.01	.43	- .90	.00
Geometry-- Application	.18	.60	.12	.85	- .64	.74	.94	1.41	1.00	1.42	1.45	2.03	.89
Graphs	- .03	.77	.60	.19	.55	- .43	.65	.07	- .01	-1.00	- .32	- .21	-1.91
Estimation	2.12	.42	1.26	.43	-1.25	- .80	- .97	2.00	1.28	- .27	- .57	1.34	.12
Sign Numbers	-1.66	3.88	-2.45	-2.23	1.26	.55	- .01	-1.07	-1.04	-1.24	.15	- .57	.39
Fractions	.85	- .99	-1.45	.79	1.68	-1.52	.83	.03	-2.44	2.70	- .03	.75	.86
Blocks	.41	.74	-1.22	.27	1.19	1.36	.80	- .49	- .76	-3.18	- .08	.06	1.35
Unclassified	- .41	-1.25	1.11	-1.40	- .78	- .53	-1.88	-1.58	-.84	-1.01	- .60	-1.85	-2.48

Table 12

Weighted Standardized Mean Residuals of Within School Item Difficulties by Content Category

Schools Within Census Division 7

Book 4, 13 year olds, 1978

	School														
Content Category	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Story Problems	-1.65	.19	1.18	.23	-1.03	1.49	- .42	- .11	- .46	.13	-1.62	- .15	- .11	- .05	- .16
Calculation	.88	1.05	1.35	.10	- .40	- .89	.68	- .08	1.33	1.61	- .19	1.45	- .45	1.36	.70
Geometry Definition	.67	- .05	- .18	- .79	.80	- .13	-2.08	1.56	- .87	-1.68	-2.67	.25	-1.07	- .32	-1.50
Geometry Application	.55	3.27	- .73	-1.20	-1.76	- .50	.88	- .42	.13	.46	- .38	2.03	- .74	2.46	.32
Graphs	-1.31	-1.91	1.23	1.33	.39	1.09	.81	1.32	1.17	- .25	-3.43	- .67	.90	-2.87	.01
Estimation	1.46	.89	- .31	.78	-1.21	.06	3.59	-1.15	.80	.33	- .00	.69	- .55	1.35	- .86
Sign Numbers	.59	-1.54	.12	- .58	2.85	-2.55	-1.94	- .38	-2.42	-2.54	-2.47	- .76	3.35	- .57	2.44
Fractions	1.03	1.10	1.40	1.58	- .91	.33	- .41	- .64	1.74	1.01	-2.80	-1.07	- .66	-1.81	.45
Blocks	- .73	.26	1.19	.19	- .75	- .41	.51	1.05	1.06	1.47	-1.80	.09	- .05	.50	.24
Unclassified	-1.19	-2.97	-1.93	- .98	1.81	.71	- .47	- .87	-1.89	.47	-3.94	-1.70	.10	- .20	- .67

Table 13

Stem and Leaf Display of p-Values for Cognitive Items--

Book 5, 13 year olds, 1978

(N = 63)

Stem	Leaf						Frequency
90	5.17	5.04					2
90	4.80	0.78					2
80	8.18	7.29	5.54				3
80	4.24	3.75	2.78	2.66	2.37	1.48	6
70	8.59	6.44	6.28				3
70	4.45	1.12	1.04	0.15			4
60	9.70	6.45					2
60	4.09	2.27	0.48				3
50	8.69	7.15	7.07	6.86			4
50	4.47	3.61	3.33	2.52	2.11	1.06	6
40	8.86	8.25	7.97	5.94			4
40	3.30	2.65	2.40	2.40	0.78		5
30	8.91	7.49	6.96	5.95			4
30	2.33						1
20	9.85	5.59	—				2
20	4.21	2.22	1.20				3
10	9.22	8.81					2
10	3.28						1
0	9.91	9.46	9.26	8.90			4
0	2.88	1.62					2
							63

Table 14

Means, Standard Deviations, Skewness, and the Pearson Product Moment Correlations Among Sato's Caution Index, Modified Caution Index and the Total Test Score for the Total Set and Two Subsets of Items

Book 5, 13 year olds, 1978

(N = 1231)

Variable	Variable 1	2	3	4	5	6	7	8	9
All Items									
1. Sato's Caution Index	1.00	.99	-.21	.66	.66	-.23	.46	.46	-.18
2. Modified Caution Index		1.00	-.15	.66	.66	-.17	.48	.48	-.13
3. Total Test Score			1.00	-.02	-.04	.96	.12	.14	.64
Arithmetic Items									
4. Sato's Caution Index				1.00	.99	-.04	.02	.03	.00
5. Modified Caution Indes					1.00	-.06	.02	.02	-.01
6. Total Test Score						1.00	.11	.13	.50
Geometry Items									
7. Sato's Caution Index							1.00	.99	-.02
8. Modified Caution Index								1.00	.01
9. Total Test Score									1.00
Mean	.31	.16	33.74	.35	.18	20.26	.21	.10	5.72
Standard Deviation	.12	.06	10.61	.19	.09	7.20	.26	.13	1.68
Skewness	.93	.80	-.11	1.22	1.21	-.23	1.86	1.93	-.02

Table 15

Pearson Product Moment Correlations* of School, Student Background, and Affective Domain Variables with Modified Caution Index and the Total Test Score for the Total Set and Two Subsets of Items

Book 5, 13 year olds, 1978

(N = 1231)

Variable	All Items		Arithmetic Items		Geometry Items	
	Modified Caution	Total Score	Modified Caution	Total Score	Modified Caution	Total Score
1. Percent of Community unskilled	-.02	-.06	-.02	-.05	-.01	-.06
2. Percent of community not employed	.03	-.21	.01	-.19	-.04	-.15
3. Percent of community on welfare	.07	-.26	.04	-.25	-.04	-.14
4. Percent of community farm workers	-.03	-.01	-.02	.00	-.03	-.02
5. Percent of community black	.08	-.36	.05	-.35	-.08	-.20
6. Percent of community professional workers	.02	.22	-.01	.20	.07	.16
7. Percent of community skilled workers	-.03	.09	.02	.09	.00	.07
8. Number of students qualified to receive Title I	.05	-.23	.04	-.22	-.06	-.12
9. School enrollment size	.02	-.12	-.01	-.12	.02	-.09
10. Sex[a]	-.16	.01	-.11	.05	-.08	-.05
11. Family owns calculator[b]	-.06	.20	-.01	.20	.02	.12
12. Type of school[c]	-.03	.15	.02	.14	.02	.11
13. Like mathematics scale	.00	.06	.04	.05	.01	.08
14. Importance of mathematics scale	-.06	.15	-.02	.14	.02	.15
15. Easiness of mathematics scale	.02	.31	.05	.29	.09	.20

[a]Sex is coded 1 for males and 2 for females.

[b]Family owns calculator is coded 0 for no and 1 for yes.

[c]Type of school is coded 1 for public and 2 for private Catholic.

*Correlations greater in absolute magnitude than .07 are significant at $p < .01$.

Table 16

Analysis of Variance Summary Results for the Modified Caution Index with the Test Score as a Covariate for the Total Set and Two Subsets of Items

Book 5, 13 year olds, 1978

(N = 1231)

		All Items		Arithmetic Items		Geometry Items	
Source	df	Mean Square	F	Mean Square	F	Mean Square	F
ANOVA							
Main Effect							
Community size	3	.005	1.30	.005	.53	.048	2.89*
Residual	1227	.004		.009		.017	
ANCOVA							
Covariate	1	.095	26.43**	.043	4.95*	.002	.11
Main Effect							
Community Size	3	.005	1.49	.002	.29	.047	2.85*
Residual	1226	.004		.009		.017	
ANOVA							
Main Effect							
Race-Ethnicity	3	.015	4.19**	.016	1.82	.066	3.99**
Residual	1227	.004		.009		.016	
ANCOVA							
Covariate	1	.095	26.43**	.043	4.96*	.002	.11
Main Effect							
Race-Ethnicity	3	.005	1.48	.009	1.03	.065	3.96**
Residual	1226	.004		.009		.017	
ANOVA							
Main Effect							
Grade Level	4	.011	3.04*	.011	1.33	.008	.51
Residual	1226	.004		.009		.017	
ANCOVA							
Covariate	1	.095	26.44**	.043	4.96*	.002	.11
Main Effect							
Grade Level	4	.005	1.48	.011	1.32	.008	.49
Residual	1225	.004		.009		.017	
ANOVA							
Main Effect							
Census Division	8	.001	.35	.001	.13	.037	2.24*
Residual	1222	.004		.009		.016	
ANCOVA							
Covariate	1	.095	26.38**	.043	4.93*	.002	.11
Main Effect							
Census Division	8	.003	.87	.002	.22	.037	2.24*
Residual	1221	.004		.009		.016	

*p < .05

**p < .01

Table 17

Percentages of Variation for the Components in Mathematics Achievement and on the Modified Caution Index

Book 5, 13 year olds, 1978

	Between Census Division	School Within Census Divisions	Error
Mathematics Achievement			
All items	5.19	23.32	71.49
Arithmetic items	4.31	22.49	73.49
Geometry items	2.01	9.78	88.21
Modified Caution Index			
All items	.00	3.51	96.49
Arithmetic items	.00	2.66	97.34
Geometry items	1.12	2.33	96.55

Table 18

Summary of F Ratios for Hierarchical Analyses of Variance for the Modified Caution Index on the Total Set and Two Subsets of Items

Book 5, 13 year olds, 1978

Effect	All Items	Arithmetic Items	Geometry Items
Census Division	.57	.33	4.37**
Schools within Census Division	1.56**	1.44**	1.39*
Schools within Census Division 1	1.51	1.52	3.92**
Schools within Census Division 2	1.78*	1.21	1.15
Schools within Census Division 3	1.53	1.61	1.75*
Schools within Census Division 4	1.39	1.18	2.01
Schools within Census Division 5	.92	1.02	.37
Schools within Census Division 6	1.07	.76	.67
Schools within Census Division 7	1.13	1.76	.84
Schools within Census Division 8	1.21	.88	2.11
Schools within Census Division 9	2.71**	2.34**	1.27

*p < .01

**p < .001

Table 19

Means and Standard Deviations on Modified Caution Indices on the Total Set and Two Subsets of Items for Schools Within Census Division 1

Book 5, 13 year olds, 1978

School Number	Sample Size	All Items		Arithmetic Items		Geometry Items	
		Mean	S.D.	Mean	S.D.	Mean	S.D.
1	15	.15	.04	.11	.08	.15	.07
2	14	.19	.07	.21	.07	.25	.25
3	12	.16	.06	.15	.08	.11	.12
4	13	.14	.06	.17	.07	.08	.14
5	20	.17	.07	.19	.08	.15	.23
6	15	.14	.06	.17	.07	.03	.06
7	14	.15	.05	.18	.09	.14	.15
8	13	.16	.05	.16	.08	.09	.09

Table 20.

Means and Standard Deviations on Modified Caution Indices on the Total Set and Two Subsets of Items for Schools Within Census Division 9

Book 5, 13 year olds, 1978

School Number	Sample Size	All Items		Arithmetic Items		Geometry Items	
		Mean	S.D.	Mean	S.D.	Mean	S.D.
1	18	.16	.04	.16	.06	.12	.11
2	13	.19	.04	.23	.06	.09	.08
3	19	.18	.06	.20	.10	.04	.15
4	17	.15	.04	.16	.07	.09	.09
5	23	.14	.05	.17	.09	.08	.11
6	15	.16	.04	.17	.11	.14	.14
7	17	.18	.08	.20	.19	.15	.10
8	18	.16	.06	.18	.08	.06	.09
9	17	.19	.05	.11	.07	.09	.13
10	17	.19	.05	.19	.09	.13	.12
11	12	.14	.05	.14	.06	.08	.10
12	13	.16	.04	.20	.10	.04	.07
13	17	.17	.05	.17	.07	.14	.14
14	11	.21	.06	.18	.08	.18	.13
15	20	.14	.06	.16	.08	.09	.09
16	12	.19	.07	.20	.08	.08	.10
17	17	.12	.07	.11	.09	.11	.12
18	14	.13	.05	.17	.07	.04	.06
19	11	.14	.04	.15	.07	.09	.12
20	15	.14	.04	.17	.06	.05	.07
21	11	.21	.07	.28	.07	.10	.09

Table 21

Weighted Standardized Mean Residuals of

Within School Item Difficulties by Content Category for Schools

Within Census Division 1

Book 5, 13 year olds, 1978

Content Category	School 1	2	3	4	5	6	7	8
Story Problems	- .03	-2.10	.50	-2.73	- .61	- .46	- .16	.65
Calculations	- .49	- .16	- .78	.30	.8?	1.69	.47	.86
Geometry	-1.32	.22	- .34	-1.03	- .03	- .23	- .04	3.16
Metric	.68	1.72	-1.21	2.76	- .29	- .53	- .19	-2.50
Graphs	-1.36	- .09	.64	.45	- .69	- .97	.03	.01
Definitions	1.51	-1.18	1.00	.32	1.61	.63	-1.89	- .59
Numeration	-1.59	2.97	-2.11	.89	1.24	-1.24	.07	- .98
Fraction to Decimals	2.81	- .85	.24	- .06	-3.09	1.70	.61	-3.35
Unclassified	.97	1.81	1.89	1.21	1.34	-1.85	1.53	- .26

Table 22

Weighted Standardized Mean Residuals of Within School Item Difficulties by Content Category for Schools Within Census Division 9

Book 5, 13 year olds, 1978

School Number	Content Category								
	Story Problems	Calculations	Geometry	Metric	Graphs	Definition	Numeration	Fractions to Decimals	Unclassified
1	1.06	.59	-1.01	- .26	.80	.91	-1.27	-2.55	.96
2	- .50	- .17	.83	-1.01	1.00	-2.55	.07	1.66	.44
3	.94	- .46	.02	1.27	-1.12	-1.19	-2.52	-1.33	1.00
4	.30	-1.28	-1.34	-1.81	.78	.57	1.13	1.06	1.81
5	- .06	- .61	.33	-2.15	1.05	- .03	.50	- .62	1.44
6	.57	-2.47	- .32	1.40	.71	- .62	.62	1.53	.55
7	.99	.02	.61	1.47	.77	.79	-2.14	-2.00	-2.09
8	- .31	.68	.98	2.03	2.09	-2.80	-1.58	-2.12	.17
9	1.36	1.69	1.57	1.08	- .92	- .82	-2.10	-2.13	-2.26
10	- .49	.88	1.90	- .37	1.49	-2.74	1.22	-2.60	- .41
11	- .37	- .14	- .94	2.08	-1.11	- .23	1.29	1.66	- .43
12	1.01	- .32	2.32	-1.82	.94	-2.07	-1.31	.36	-1.23
13	- .48	.31	- .81	1.11	.29	- .39	- .31	.30	.54
14	.08	2.06	1.24	1.39	-1.16	- .87	- .01	-2.78	-1.15
15	- .23	1.05	.65	.61	-1.35	.39	.44	-1.54	- .39
16	-1.38	2.48	1.35	- .06	- .76	- .84	.35	- .21	-2.88
17	- .29	.08	-1.36	-2.89	.87	- .17	2.35	.96	.90
18	- .43	-1.22	.86	.63	.71	1.20	-1.14	- .68	- .08
19	- .10	2.15	.19	.85	.67	-1.76	- .84	- .54	-2.49
20	1.57	-1.07	1.39	- .70	1.11	-2.11	-2.20	2.02	-1.60
21	1.70	.23	-1.21	.81	.76	-2.01	2.06	-1.49	- .21

Figures 1 - 22

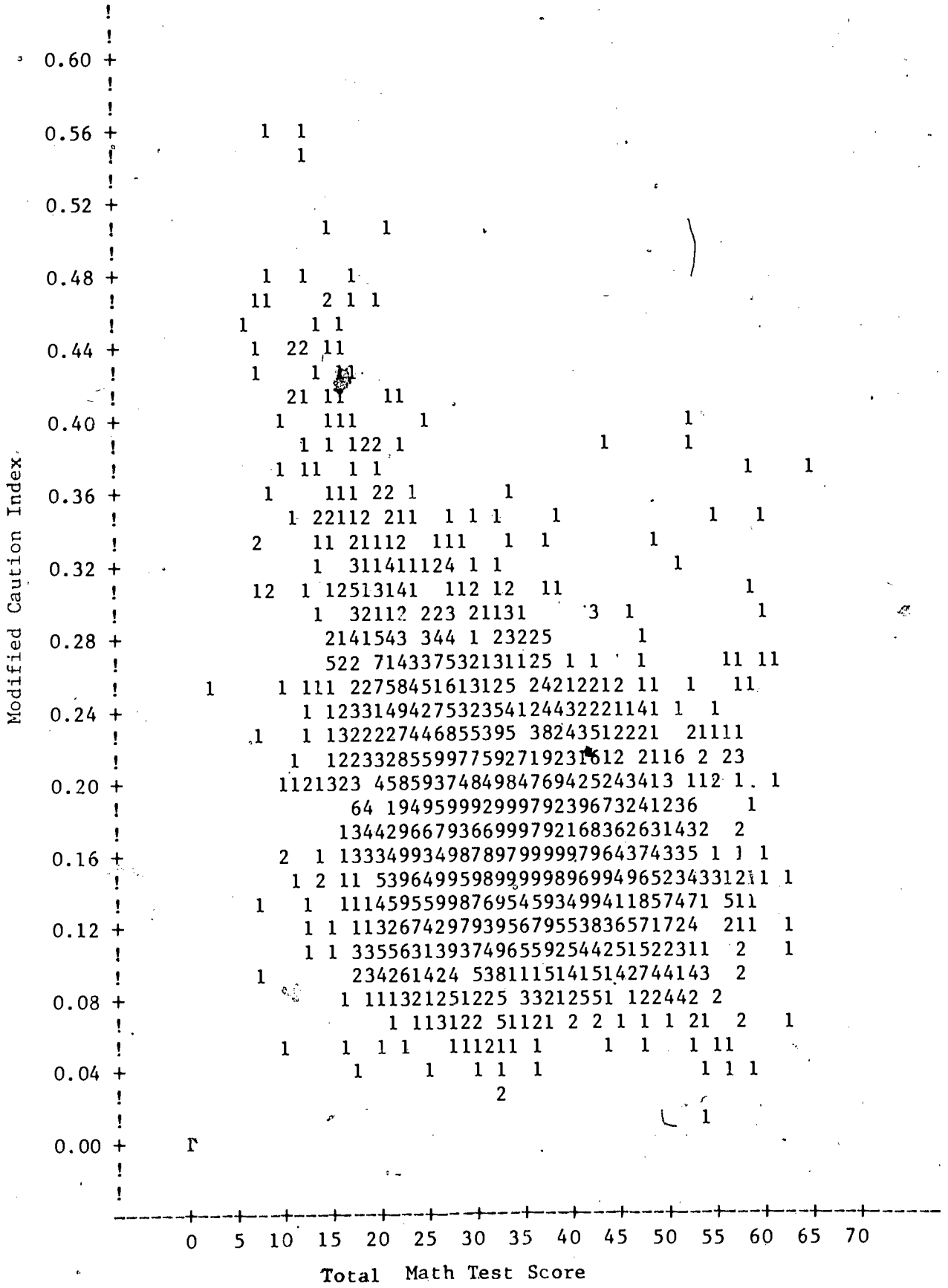

Figure 1. Scatter diagram of modified caution index with total math test score. (A 9 denotes 9 or more observations, otherwise the numeral corresponds to the number of observations; Book 4, 13 year olds, 1978.)

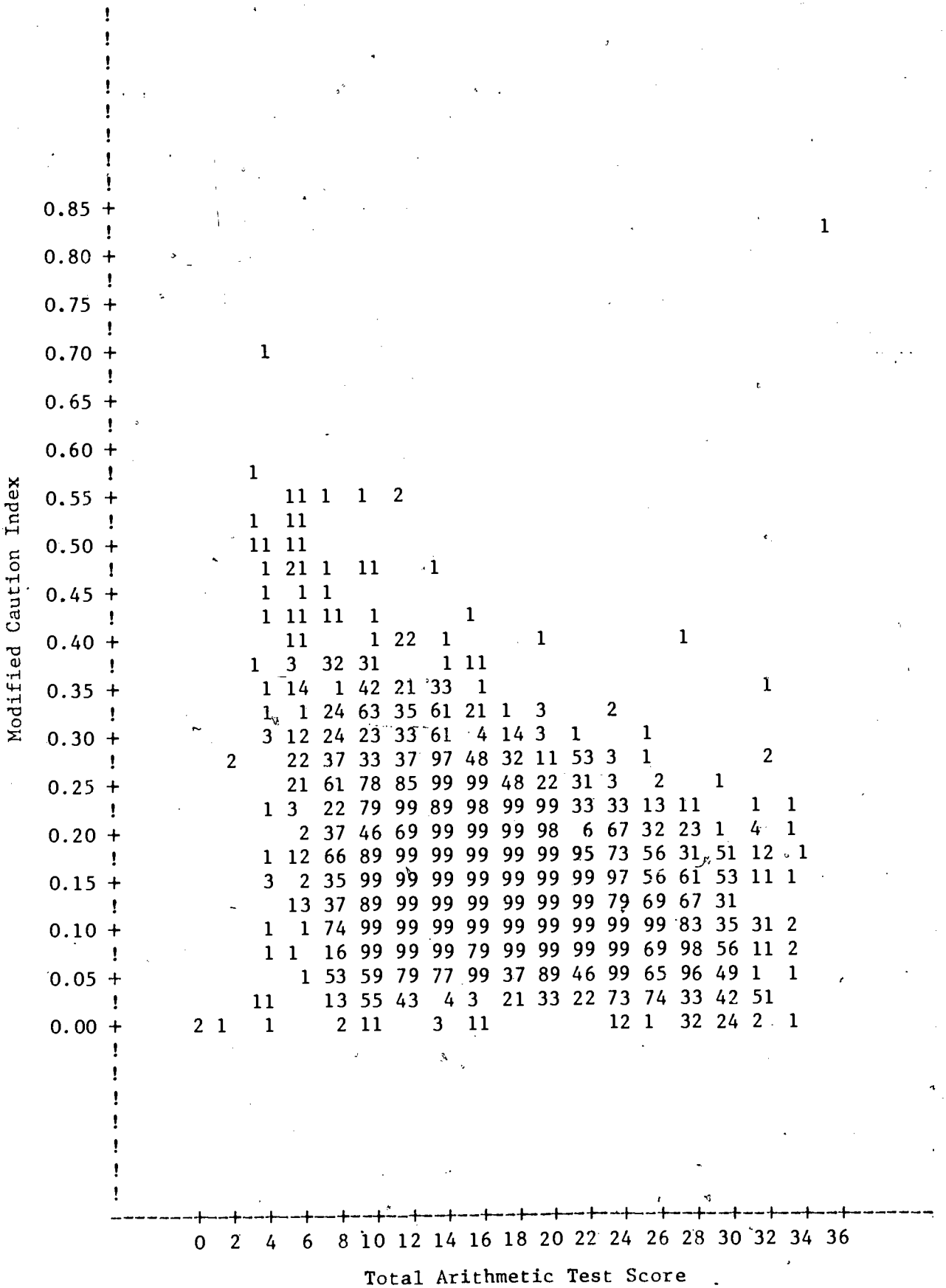

Figure 2. Scatter diagram of modified caution index with total arithmetic test score. (A 9 denotes 9 or more observations, otherwise the numeral corresponds to the number of observations; Book 4, 13 year olds, 1978.)

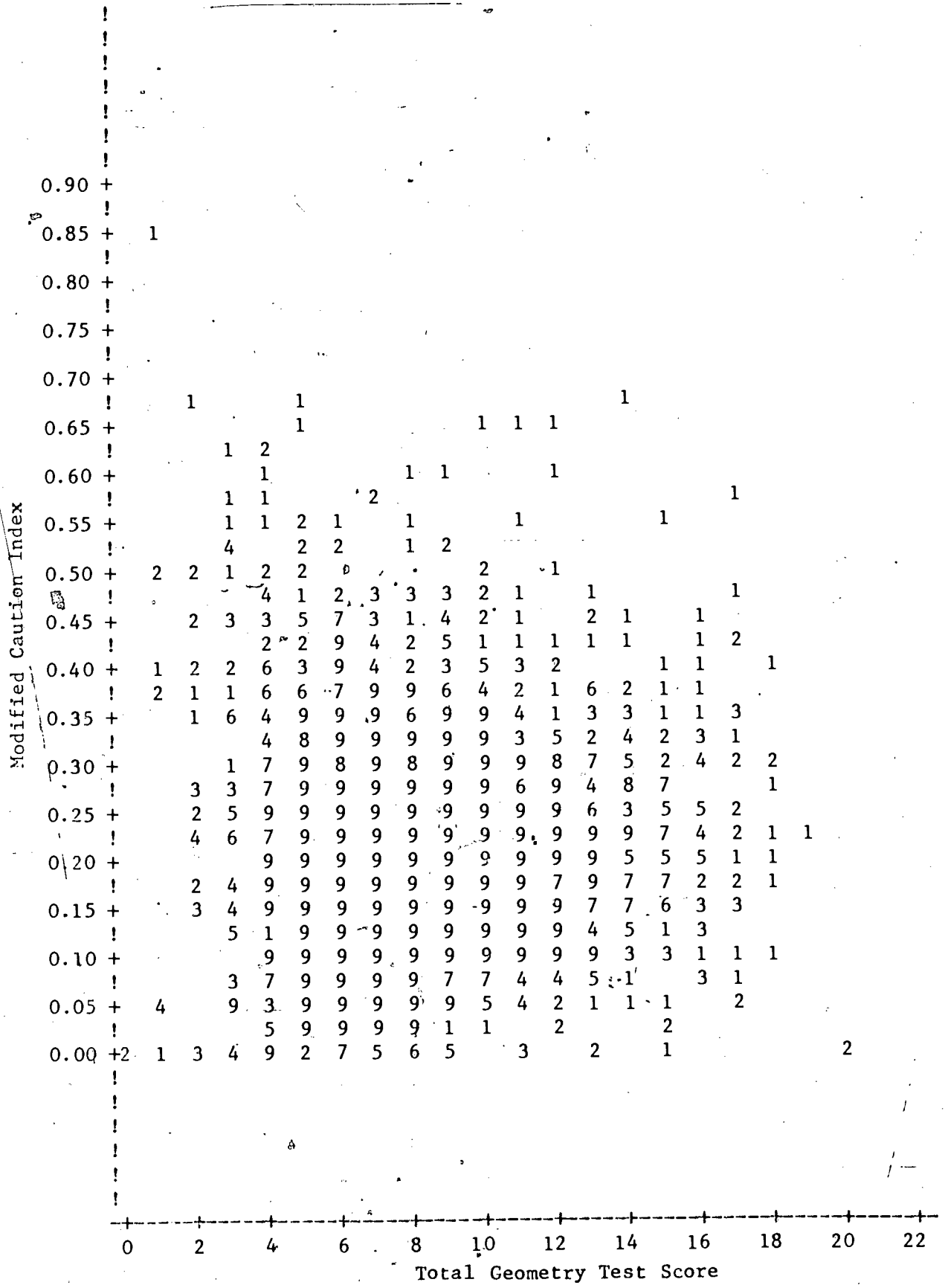

Figure 3. Scatter diagram of modified caution index with total geometry test score. (A 9 denotes 9 or more observations, otherwise the numeral corresponds to the number of observations; Book 4, 13 year olds, 1978.)

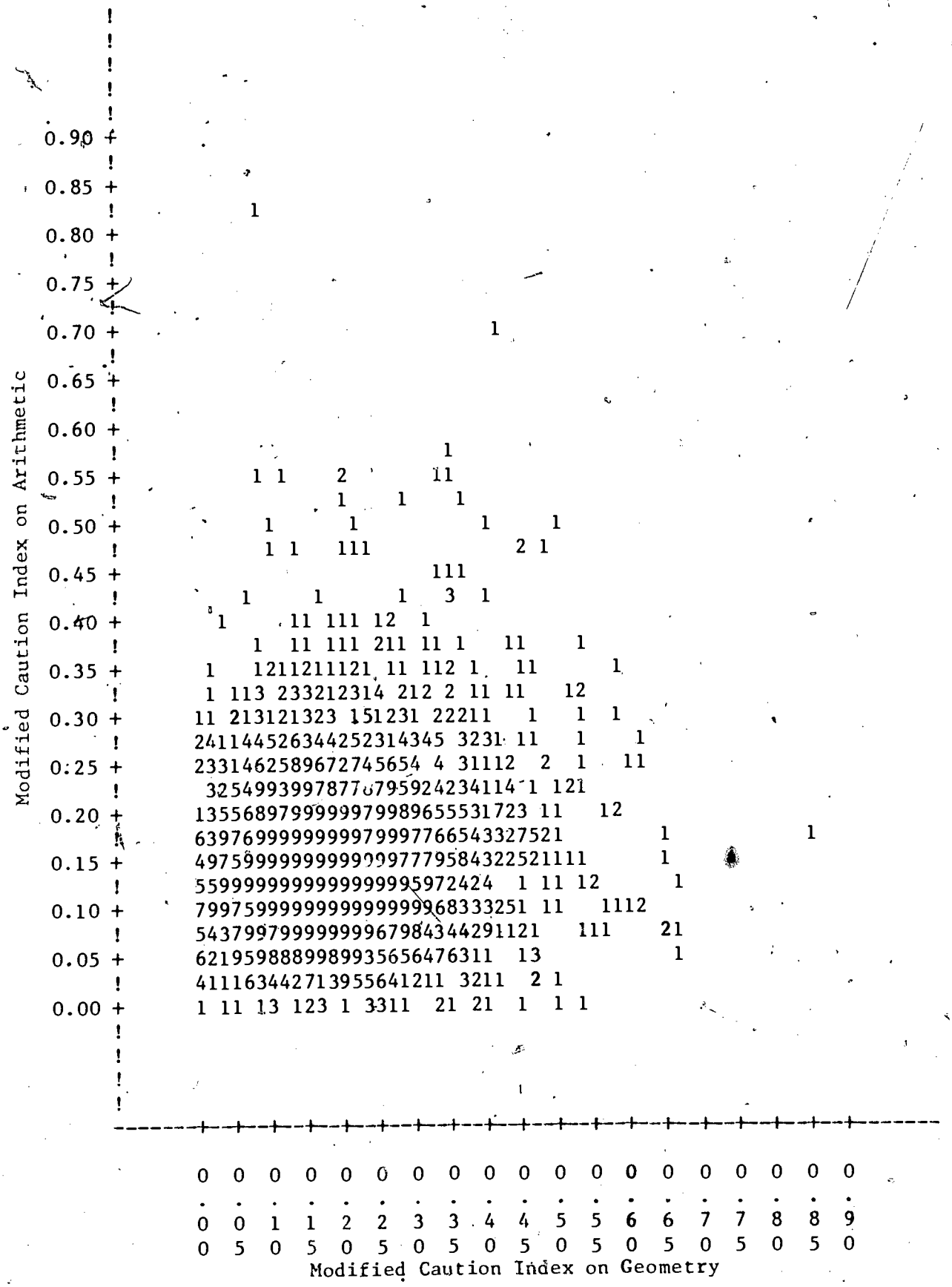

Figure 4. Scatter diagram of modified caution index from the arithmetic test with modified caution index from the geometry test. (A 9 denotes 9 or more observations, otherwise the numeral corresponds to the number of observations; Book 4, 13 year olds, 1978.)

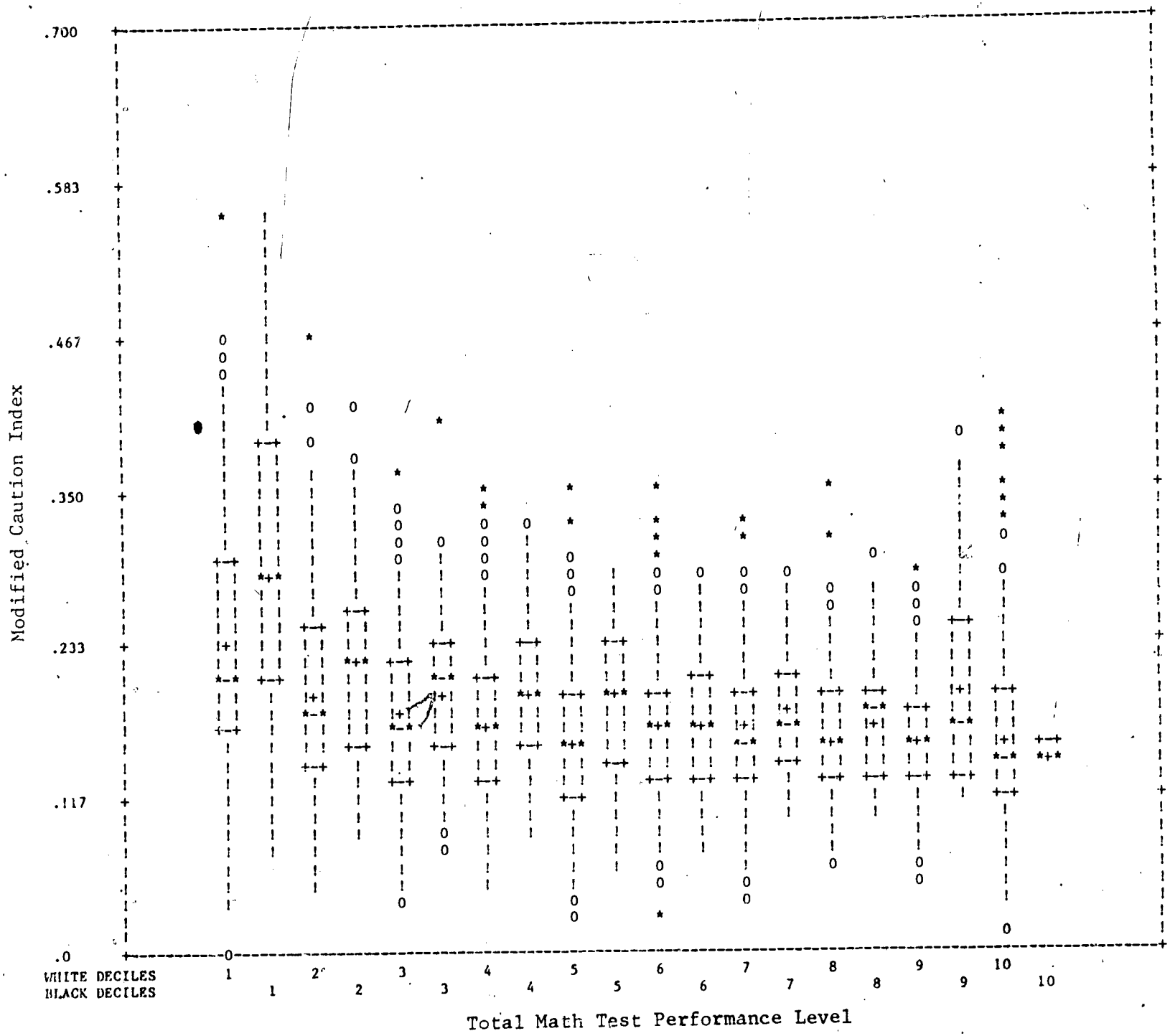

Figure 5. Box and whisker plot of modified caution index for whites and blacks with the total math test score interval split into deciles. (Book 4, 13 year olds, 1978.)

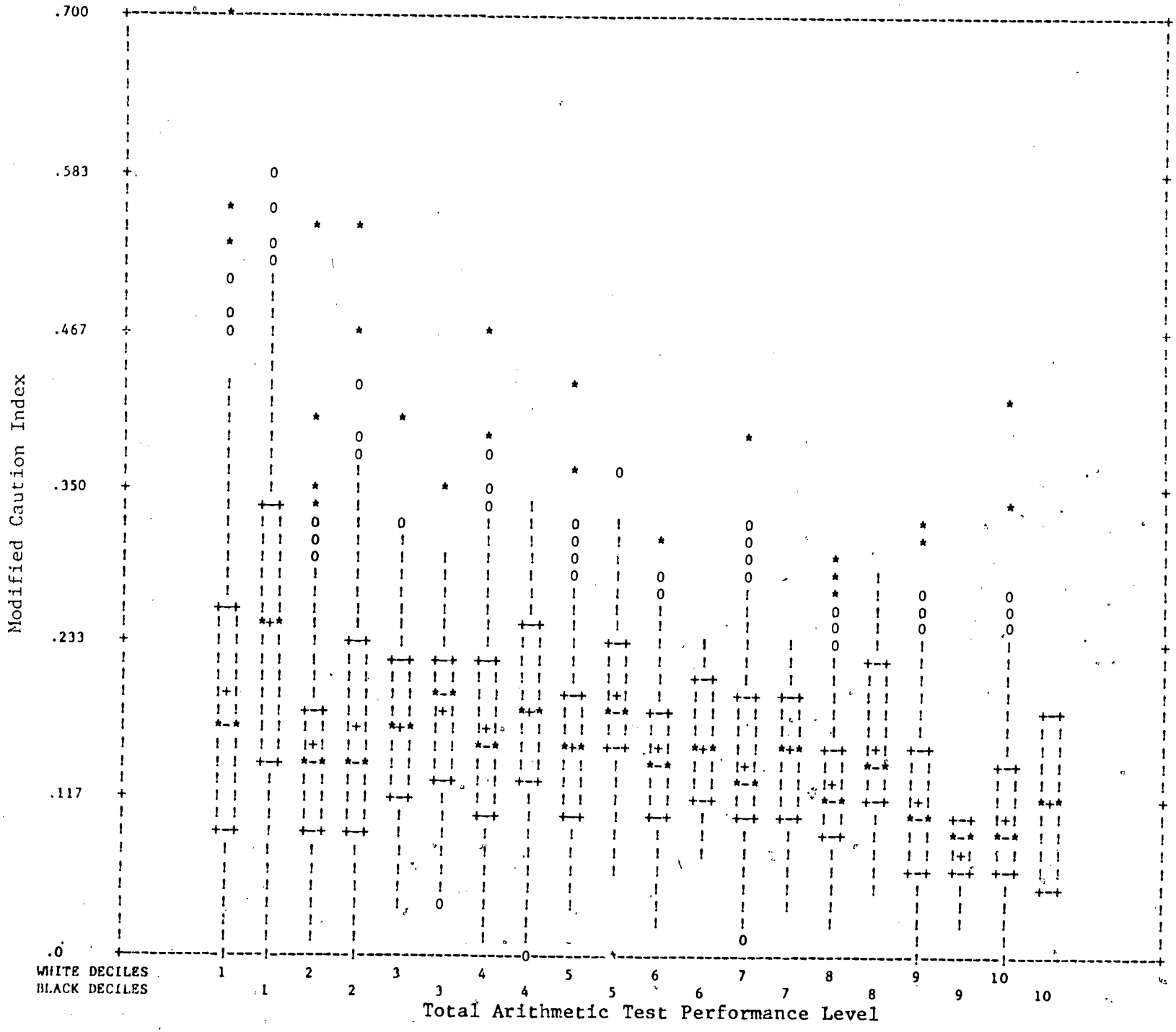

Figure 6. Box and whisker plot of modified caution index for whites and blacks with the total arithmetic test score interval split into deciles. (Book 4, 13 year olds, 1978.)

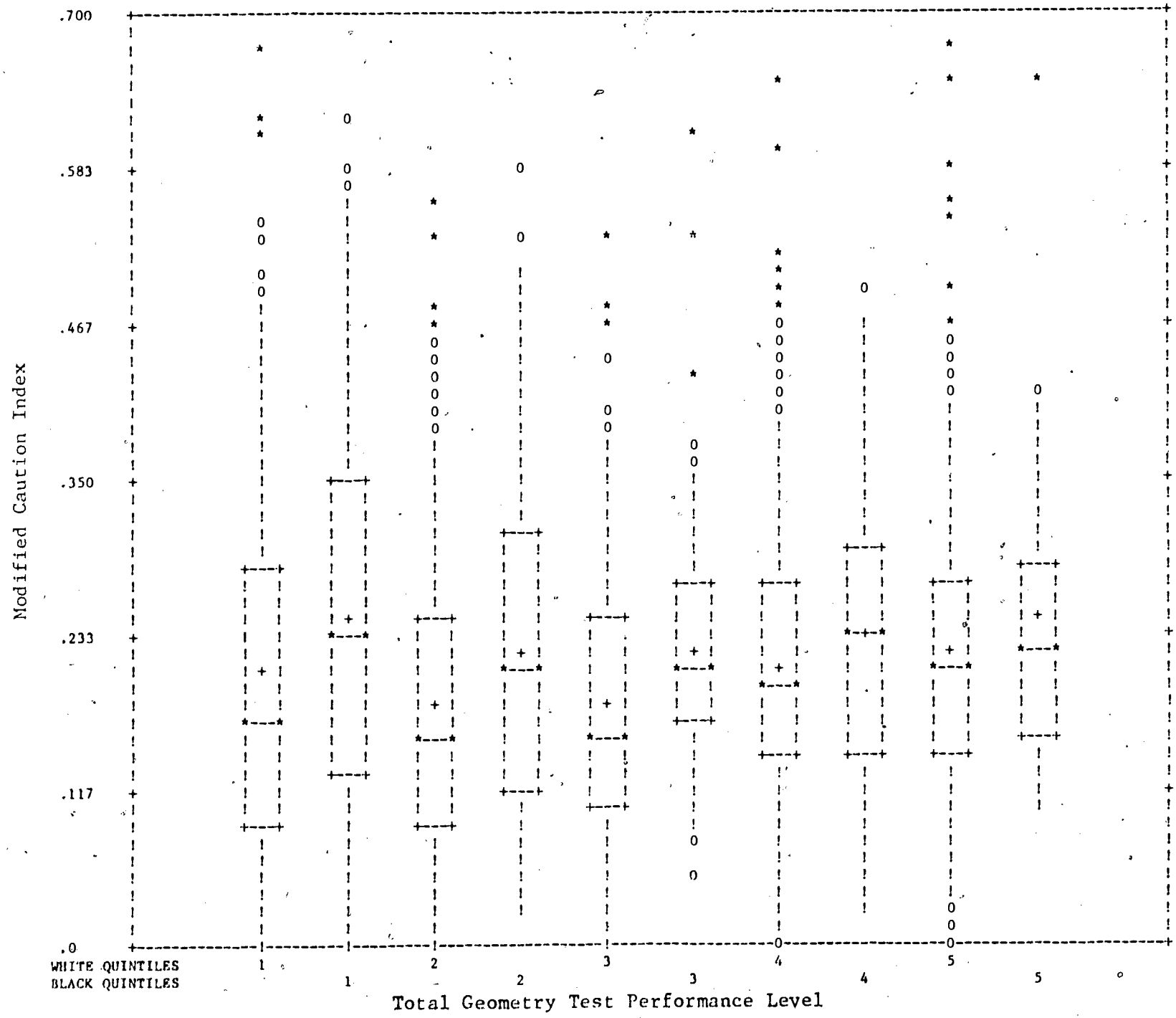

Figure 7. Box and whisker plot of modified caution index for whites and blacks with the total geometry test score interval split into quintiles. (Book 4, 13 year olds, 1978.)

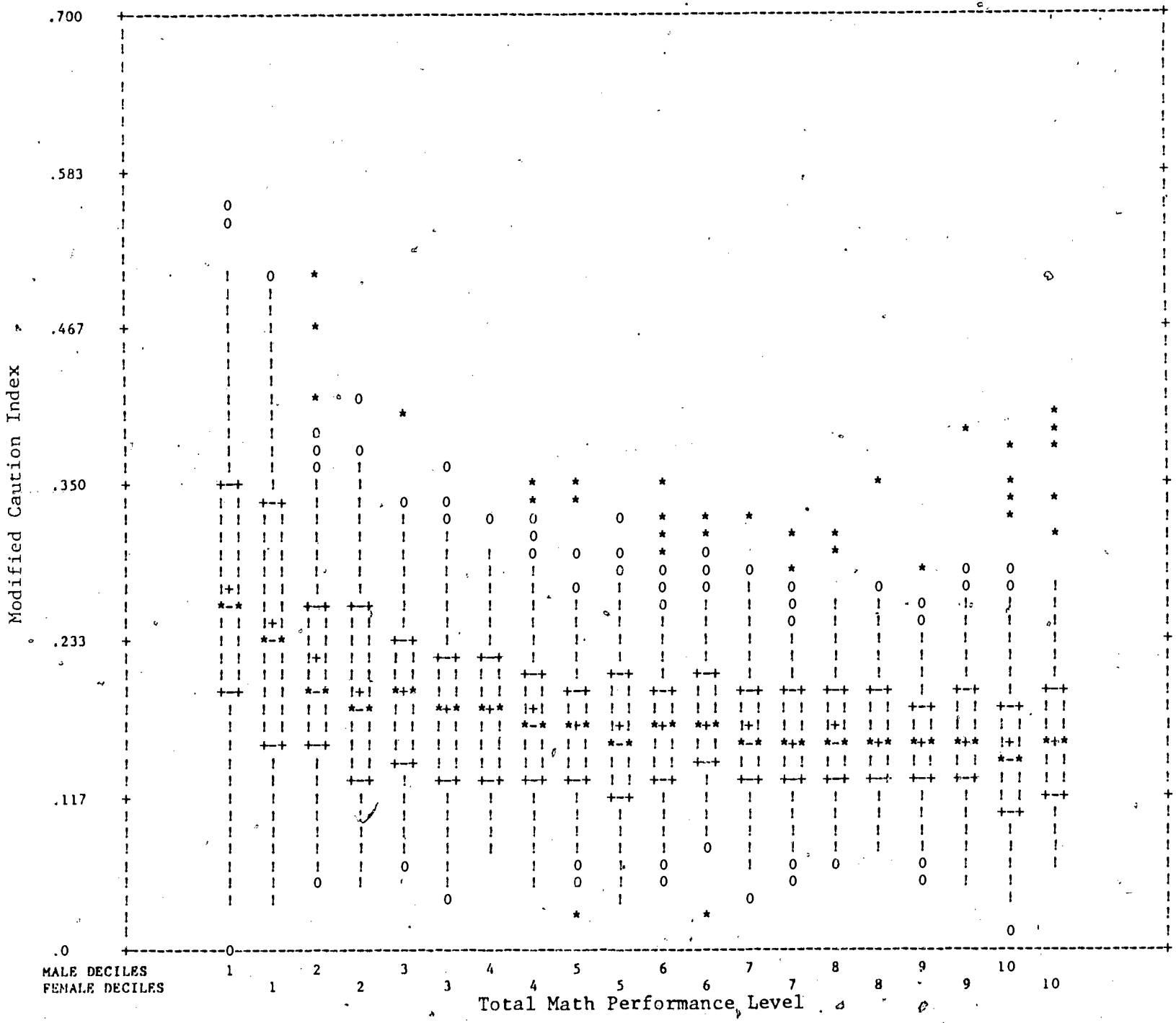

Figure 8. Box and whisker plot of modified caution index for males and females with the total math test score interval split into deciles. (Book 4, 13 year olds, 1978.)

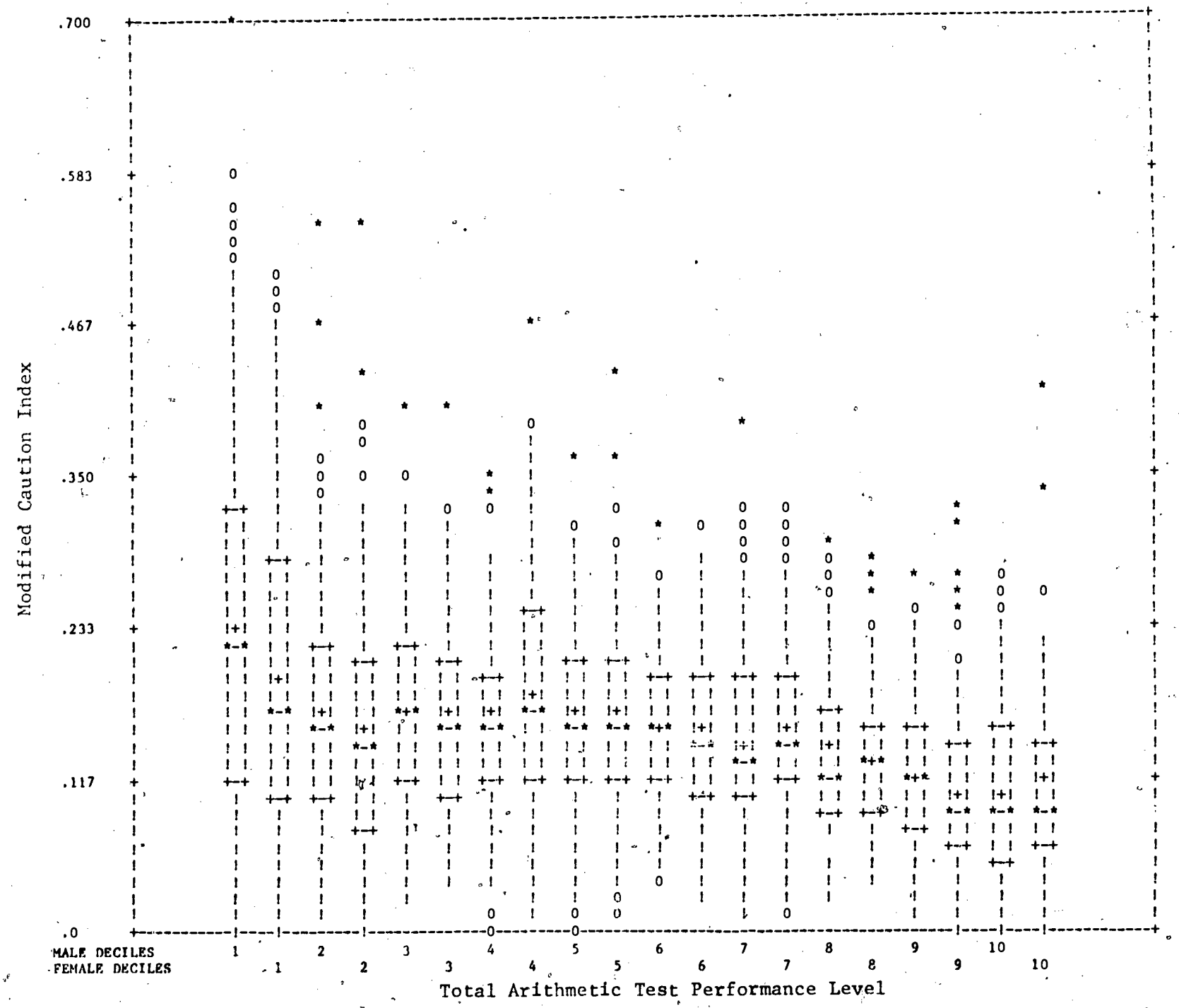

Figure 9. Box and whisker plot of modified caution index for males and females with the total arithmetic test score interval split into deciles. (Book 4, 13 year olds, 1978.)

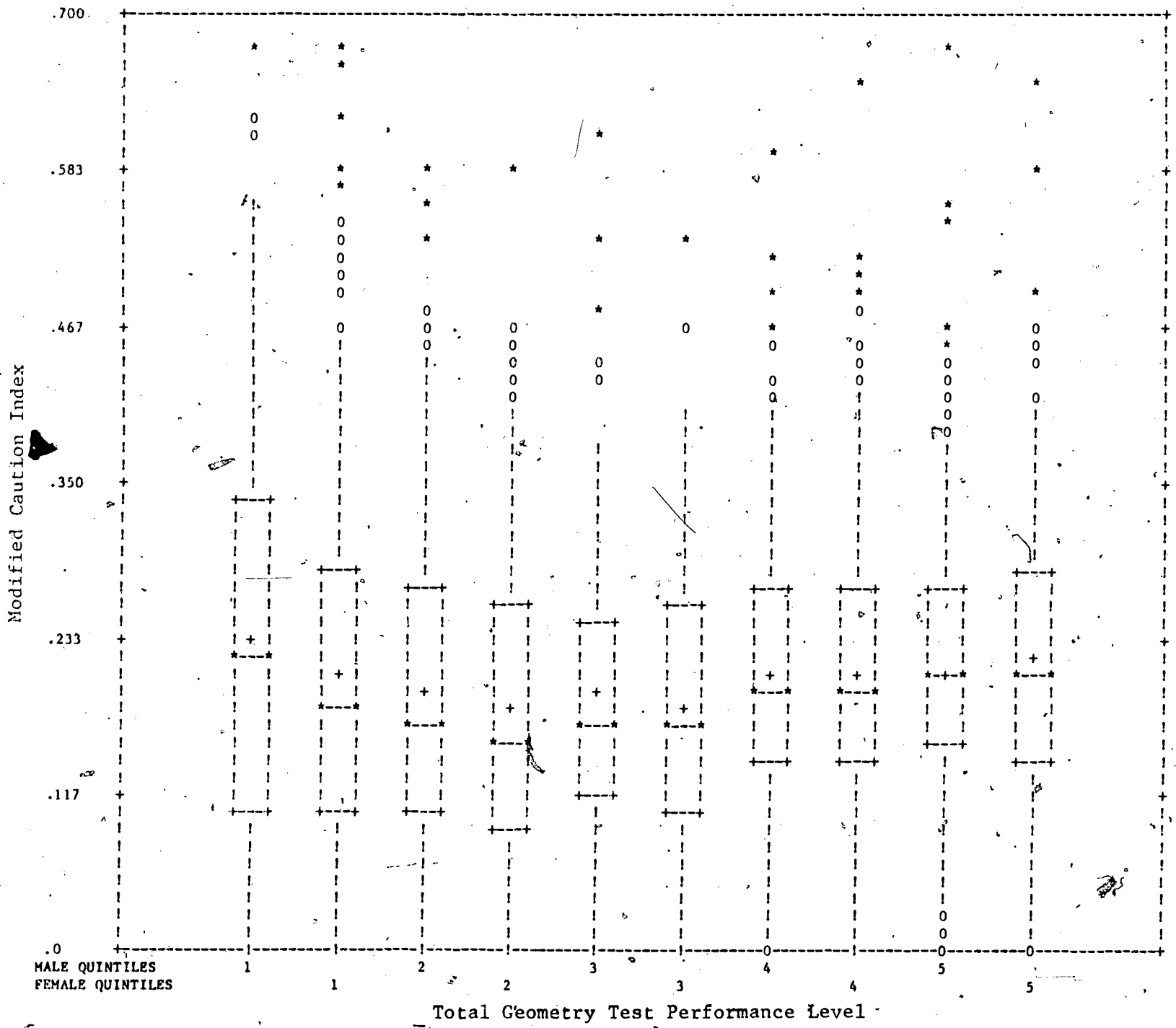

Figure 10. Box and whisker plot of modified caution index for males and females with the total geometry test score interval split into quantiles. (Book 4, 13 year olds, 1978.)

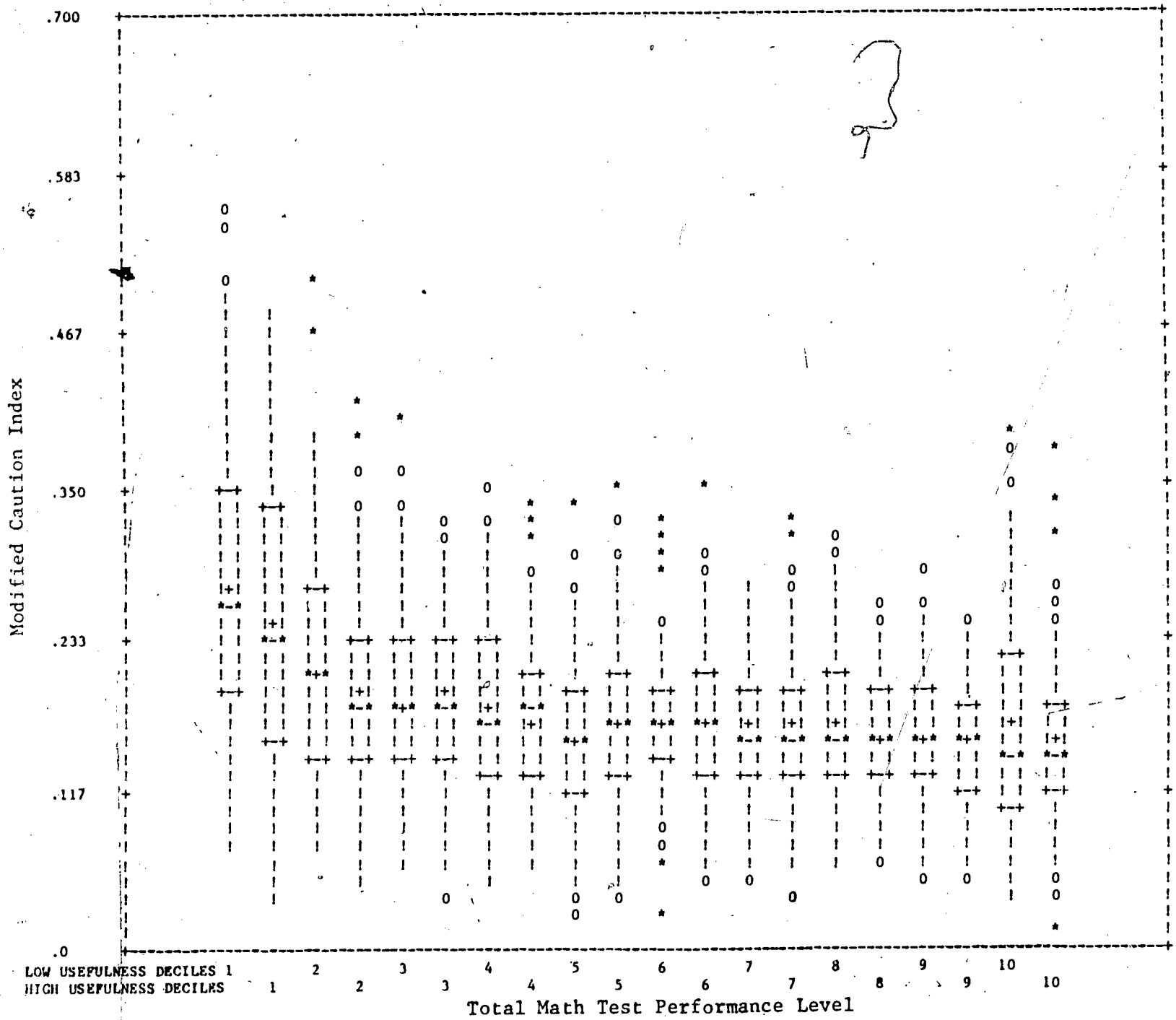

Figure 11. Box and whisker plot of modified caution index for students with low and high scores on the usefulness of mathematics scale with the total math test score interval split into deciles. (Book 4, 13 year olds, 1978.)

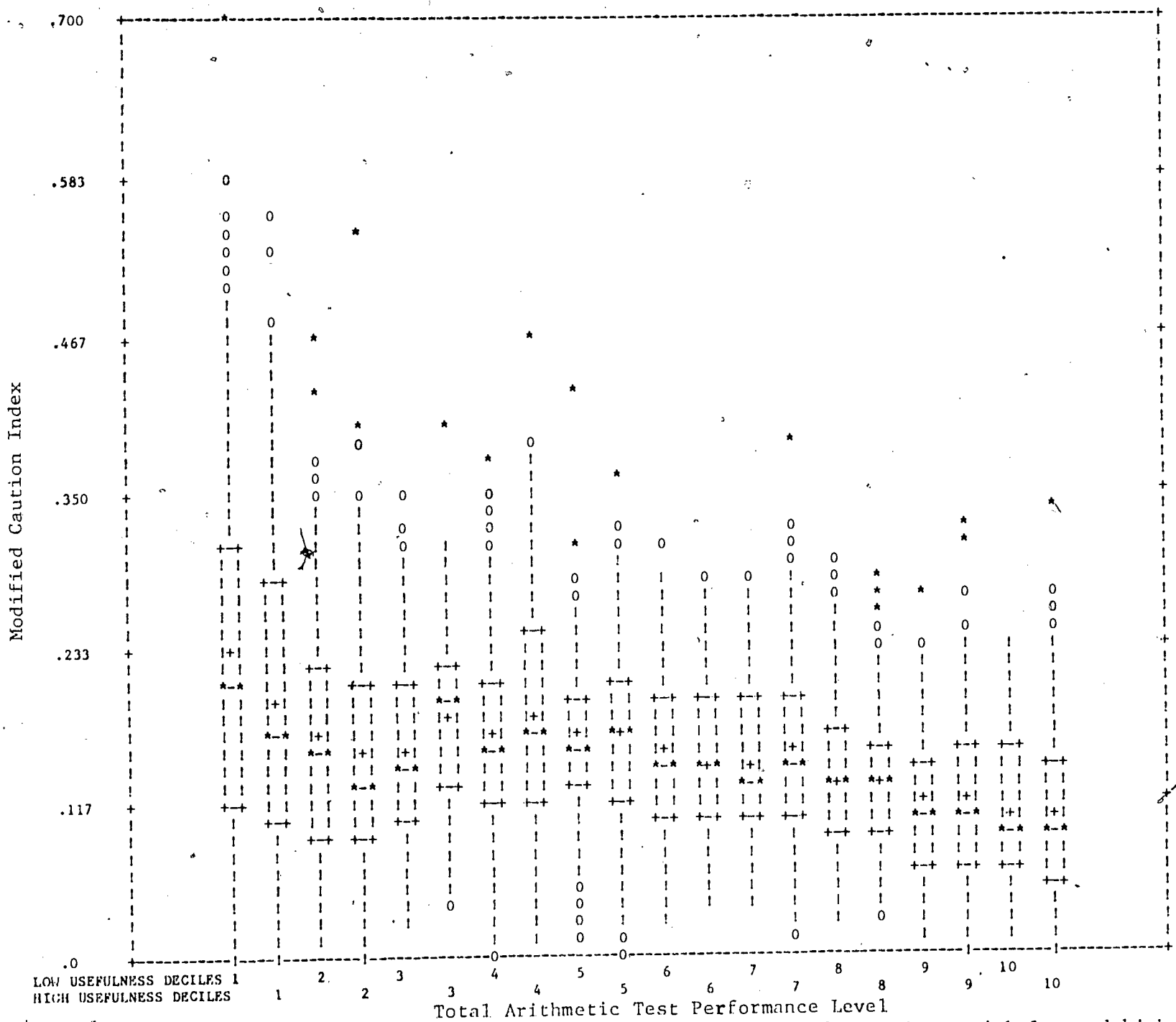

Figure 12. Box and whisker plot of modified caution index for students with low and high scores on the usefulness of mathematics scale with the total arithmetic test score interval split into deciles. (Book 4, 13 year olds, 1978.)

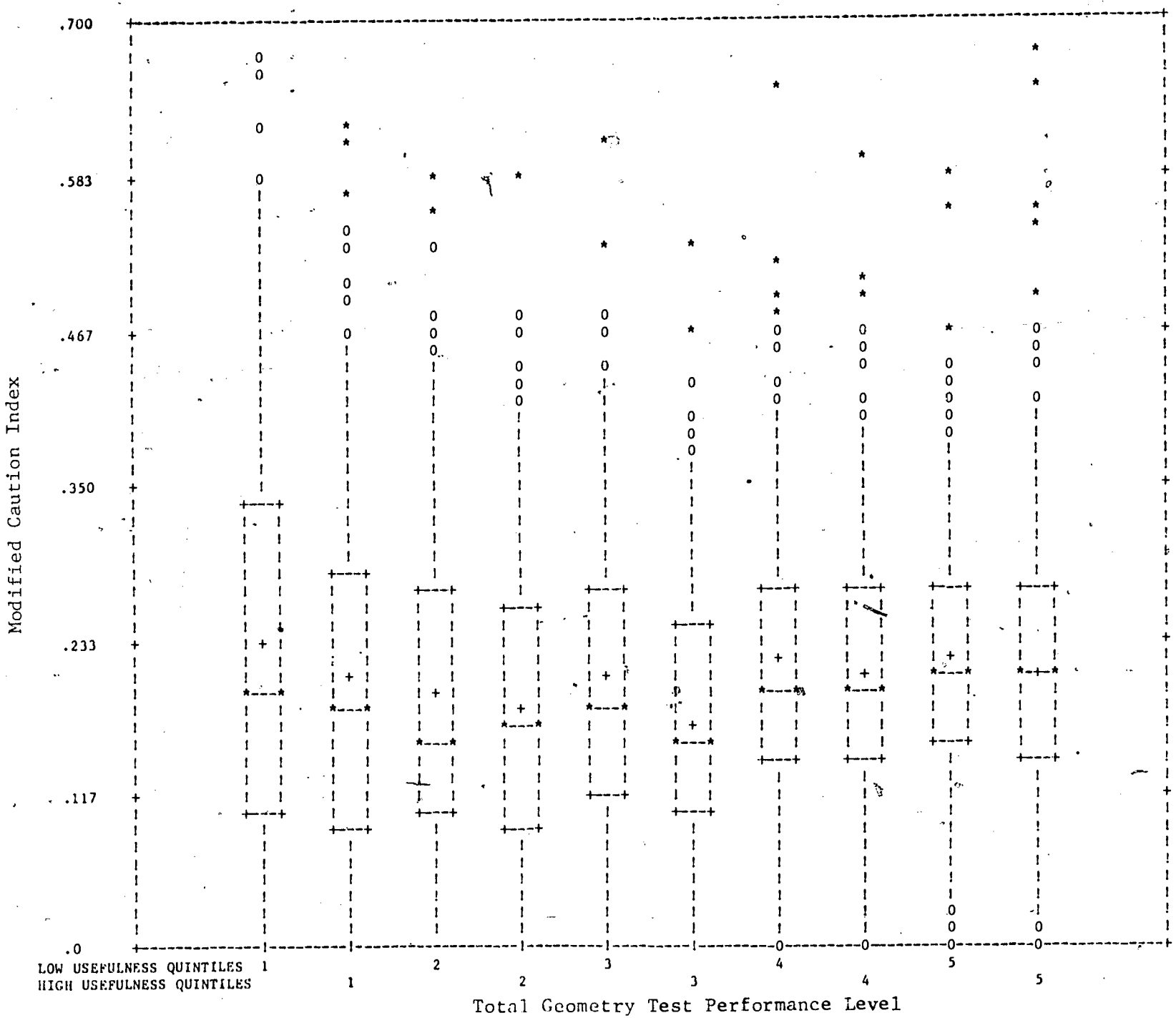

Figure 13. Box and whisker plot of modified caution index for students with low and high scores on the usefulness of mathematics scale with the total geometry test score interval split into quintiles. (Book 4, 13 year olds, 1978.)

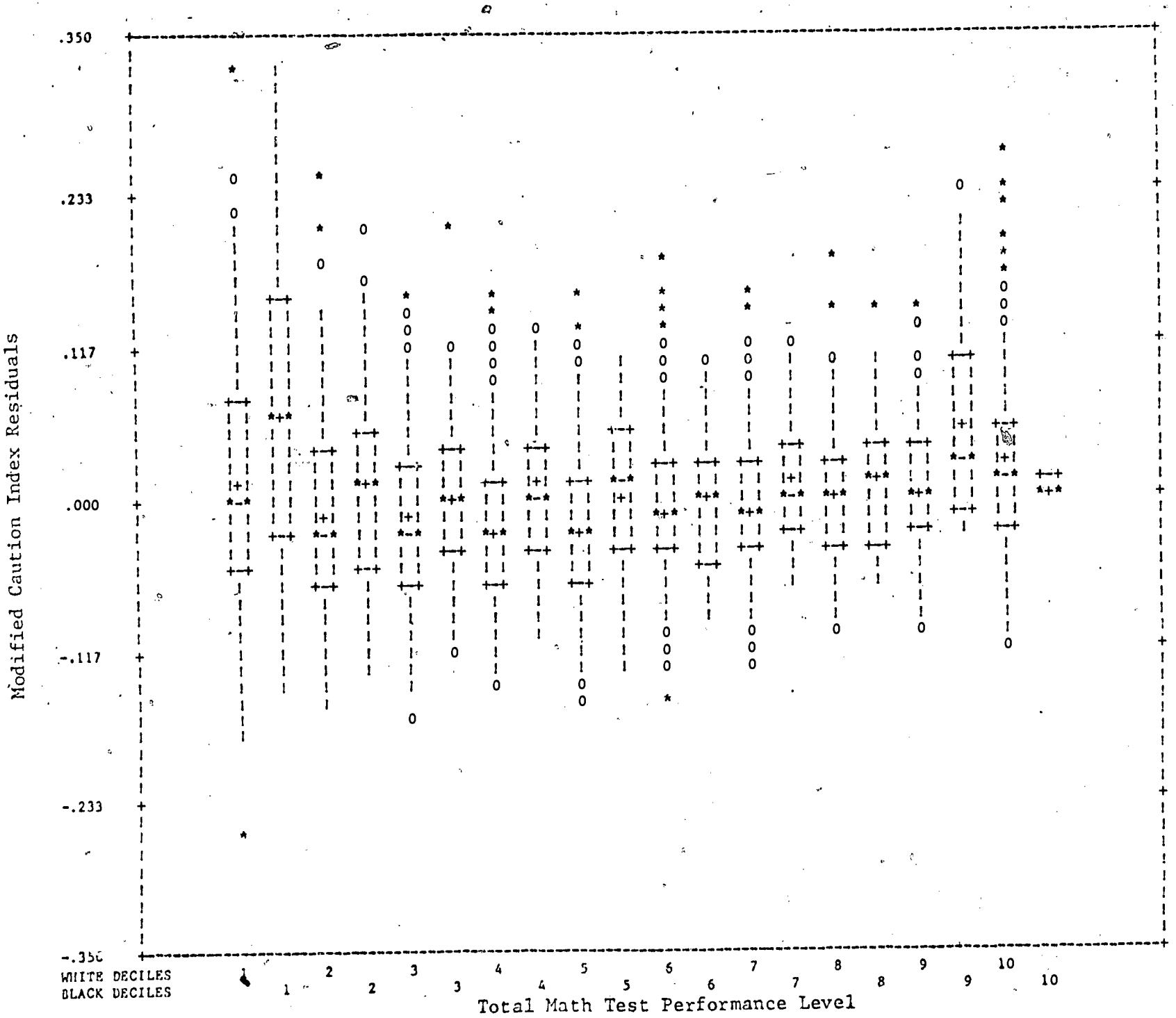

Figure 14. Box and whisker plot of modified caution index residuals for whites and blacks with the total math test score interval split into deciles. (Book 4, 13 year olds, 1978.)

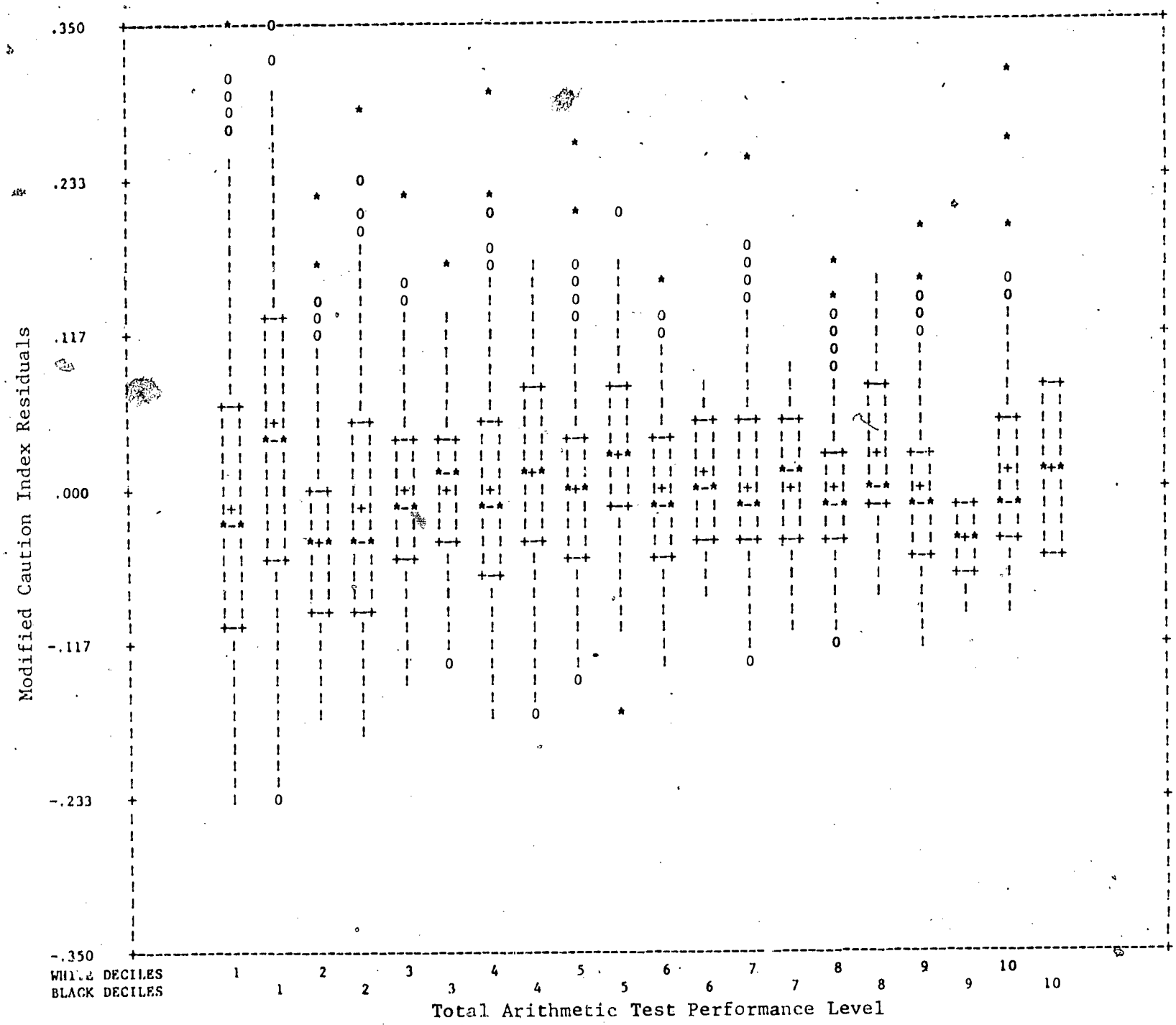

Figure 15. Box and whisker plot of modified caution index residuals for whites and blacks with the total arithmetic test score interval split into deciles. (Book 4, 13 year olds, 1978.)

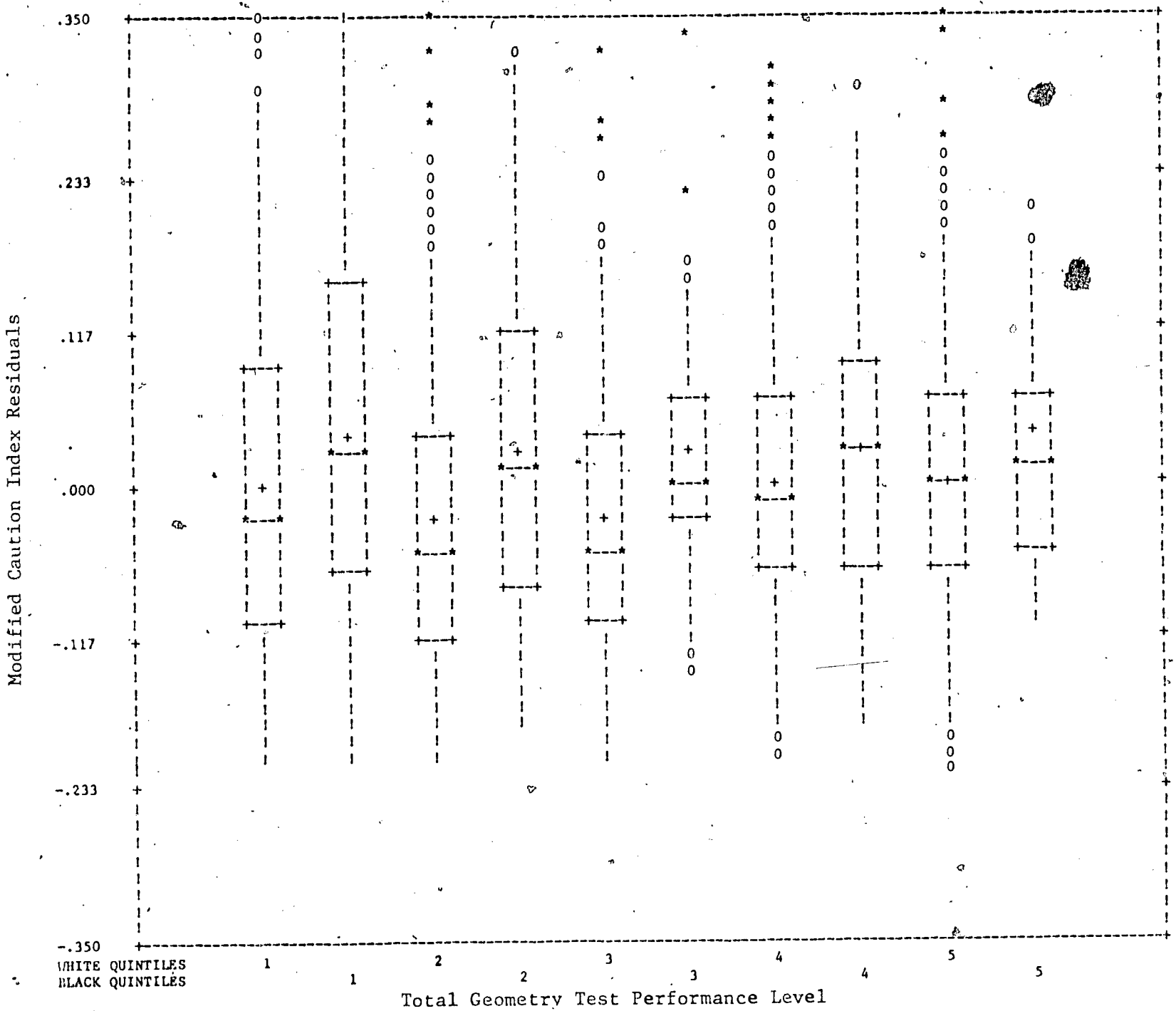

Figure 16. Box and whisker plot of modified caution index residuals for whites and blacks with the total geometry test score interval split into quintiles. (Book 4, 13 year olds, 1978.)

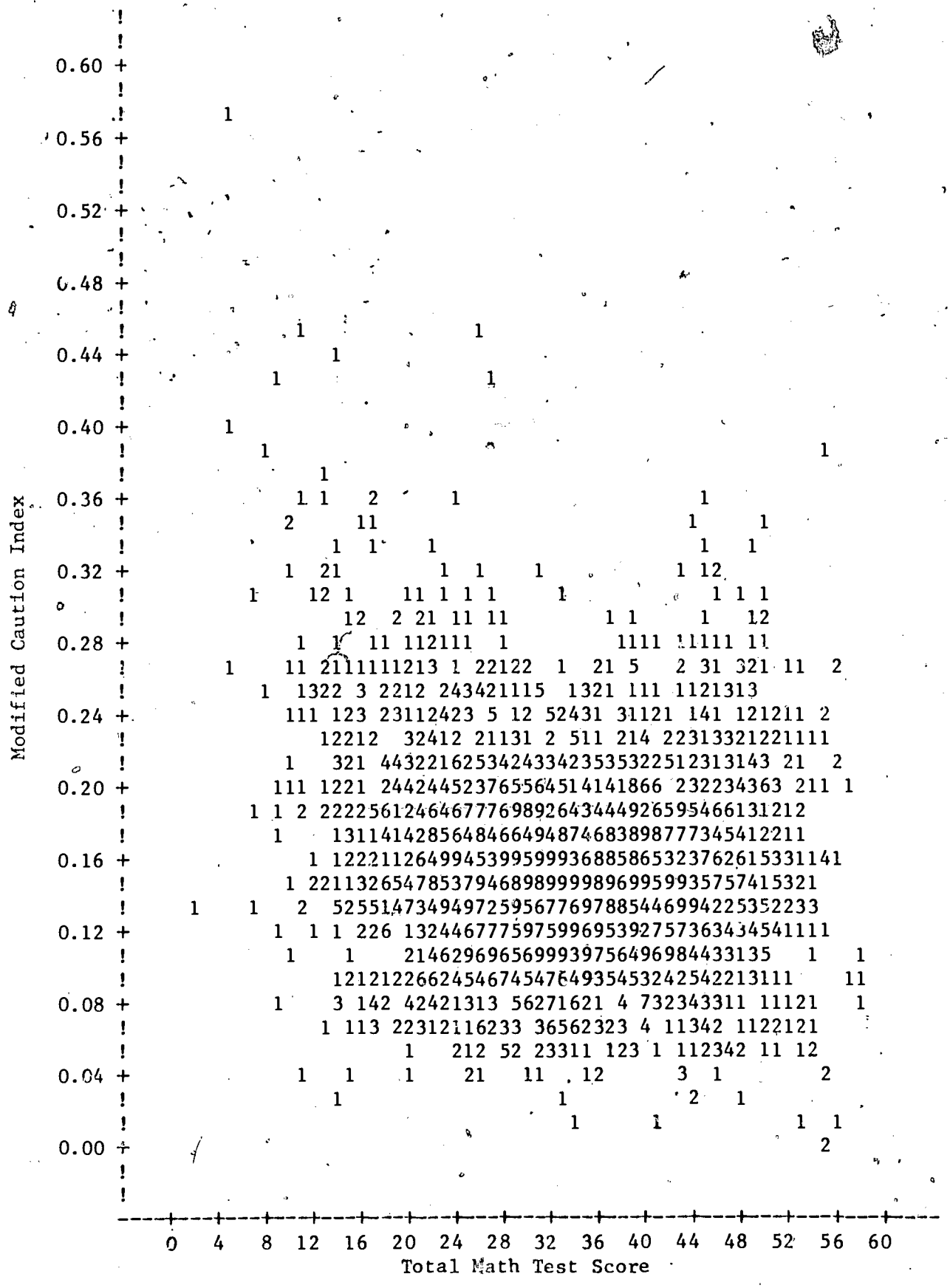

Figure 17. Scatter diagram of modified caution index with total math test score. (A 9 denotes 9 or more observations, otherwise the numeral corresponds to the number of observations; Book 5, 13 year olds, 1978.)

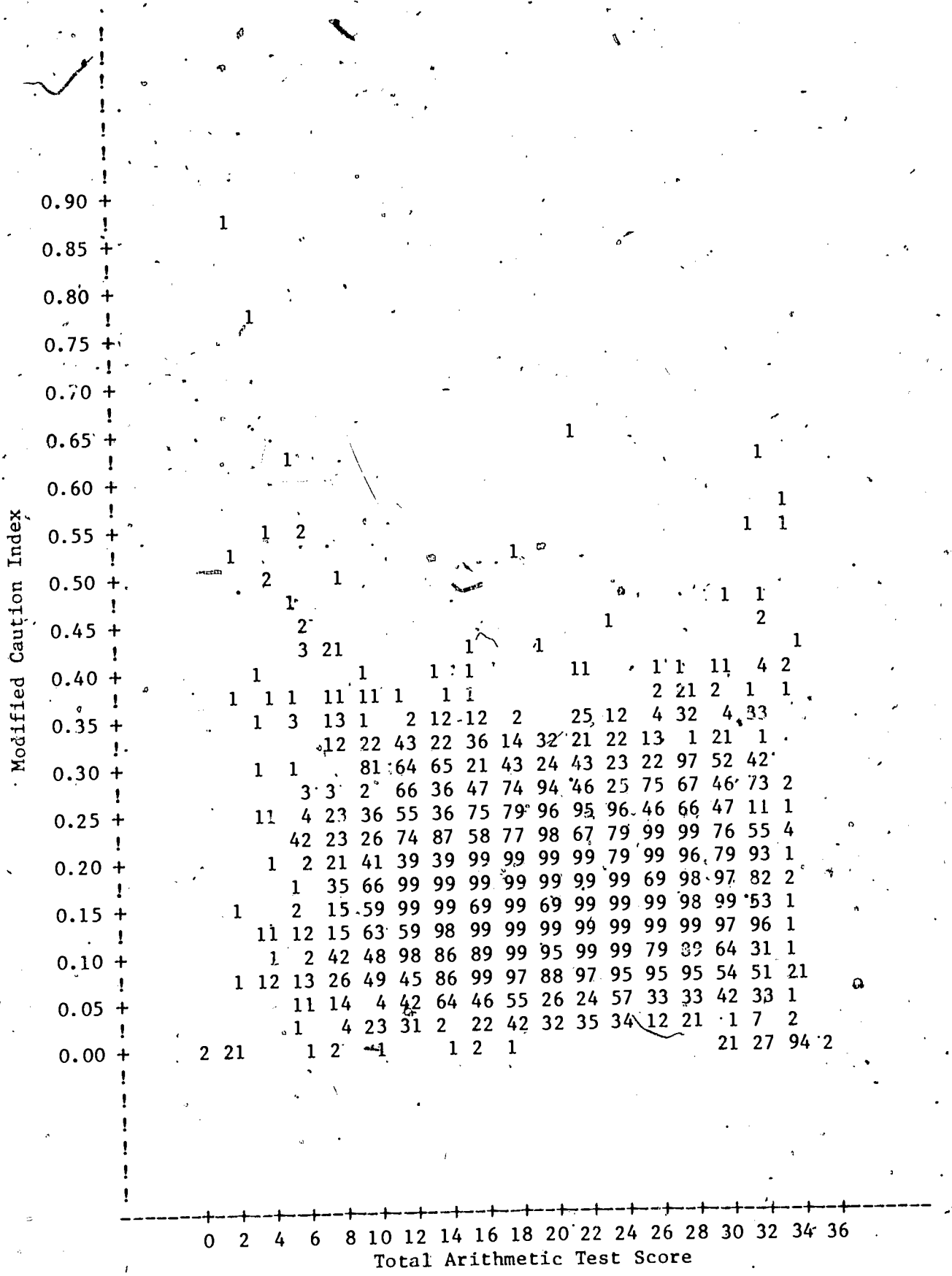

Figure 18. Scatter diagram of modified caution index with total arithmetic test score. (A 9 denotes 9 or more observations, otherwise the numeral corresponds to the number of observations; Book 5, 13 year olds, 1978.)

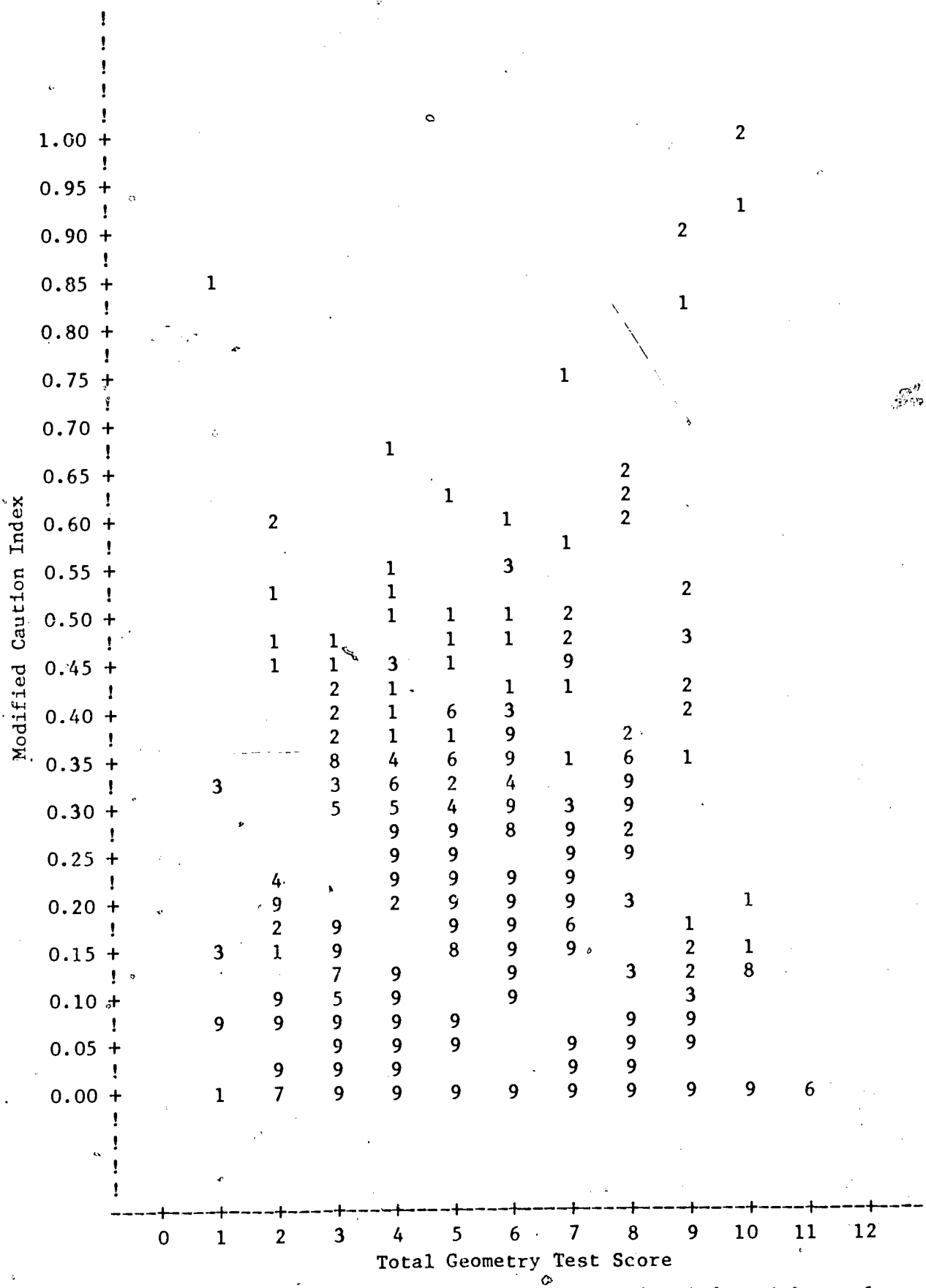

Figure 19. Scatter diagram of modified caution index with total geometry test score. (A 9 denotes 9 or more observations, otherwise the numeral corresponds to the number of observations; Book 5, 13 year olds, 1978.)

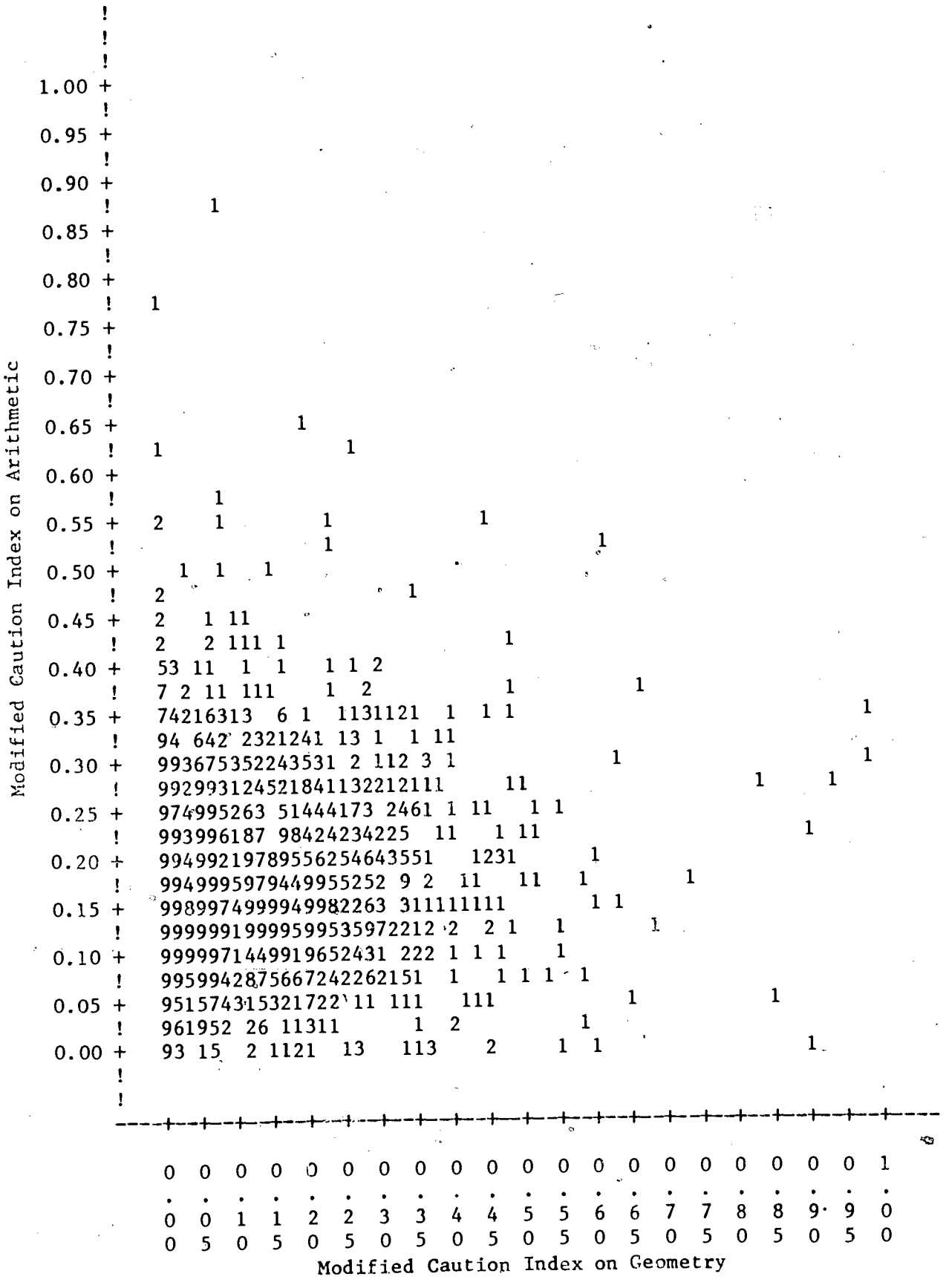

Figure 20. Scatter diagram of modified caution index from the arithmetic test with modified caution index from the geometry test. (A 9 denotes 9 or more observations, otherwise the numeral corresponds to the number of observations; Book 5, 13 year olds, 1978.)

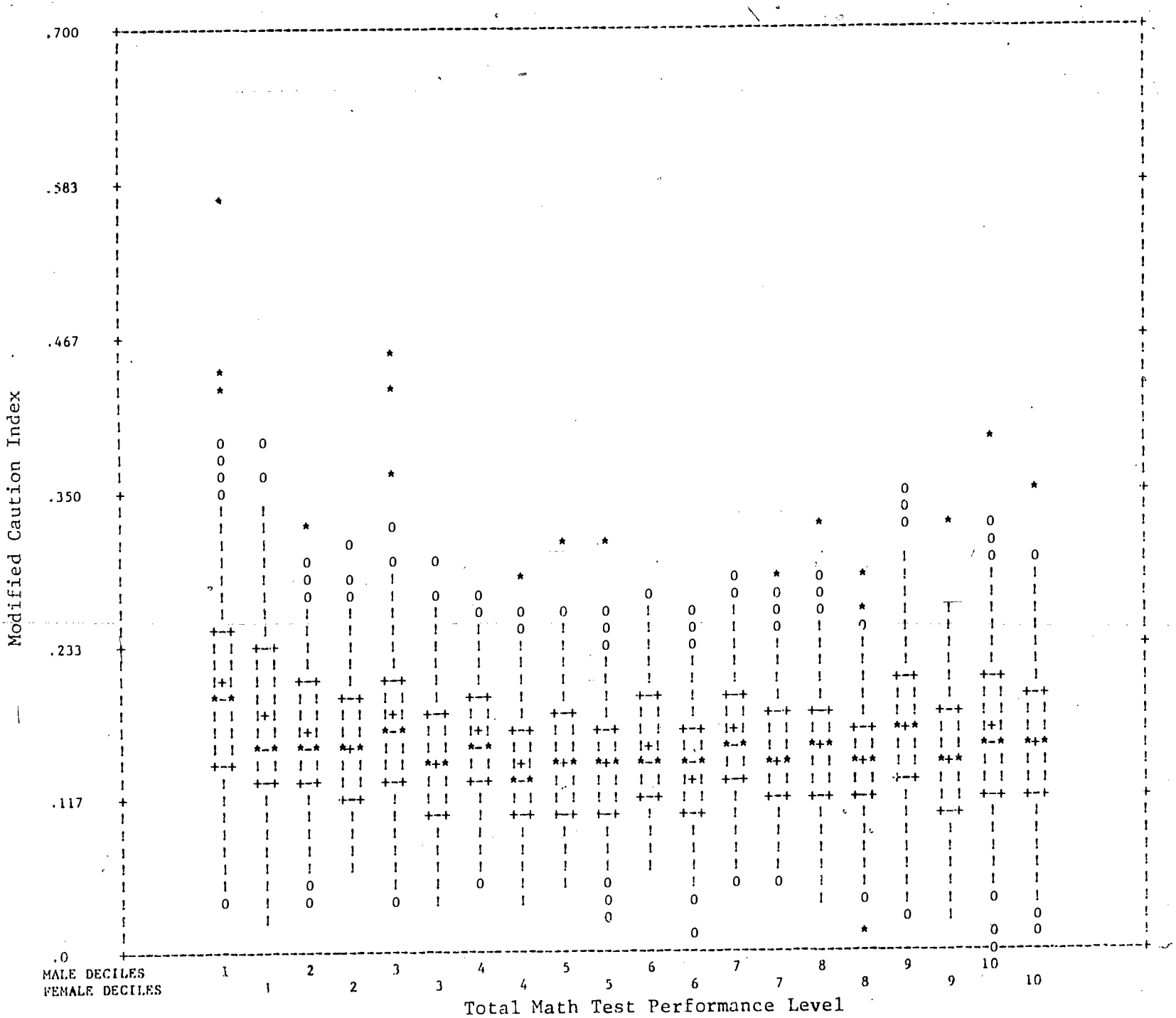

Figure 21. Box and whisker plot of modified caution index for males and females with the total math test score interval split into deciles. (Book 5, 13 year olds, 1978.)

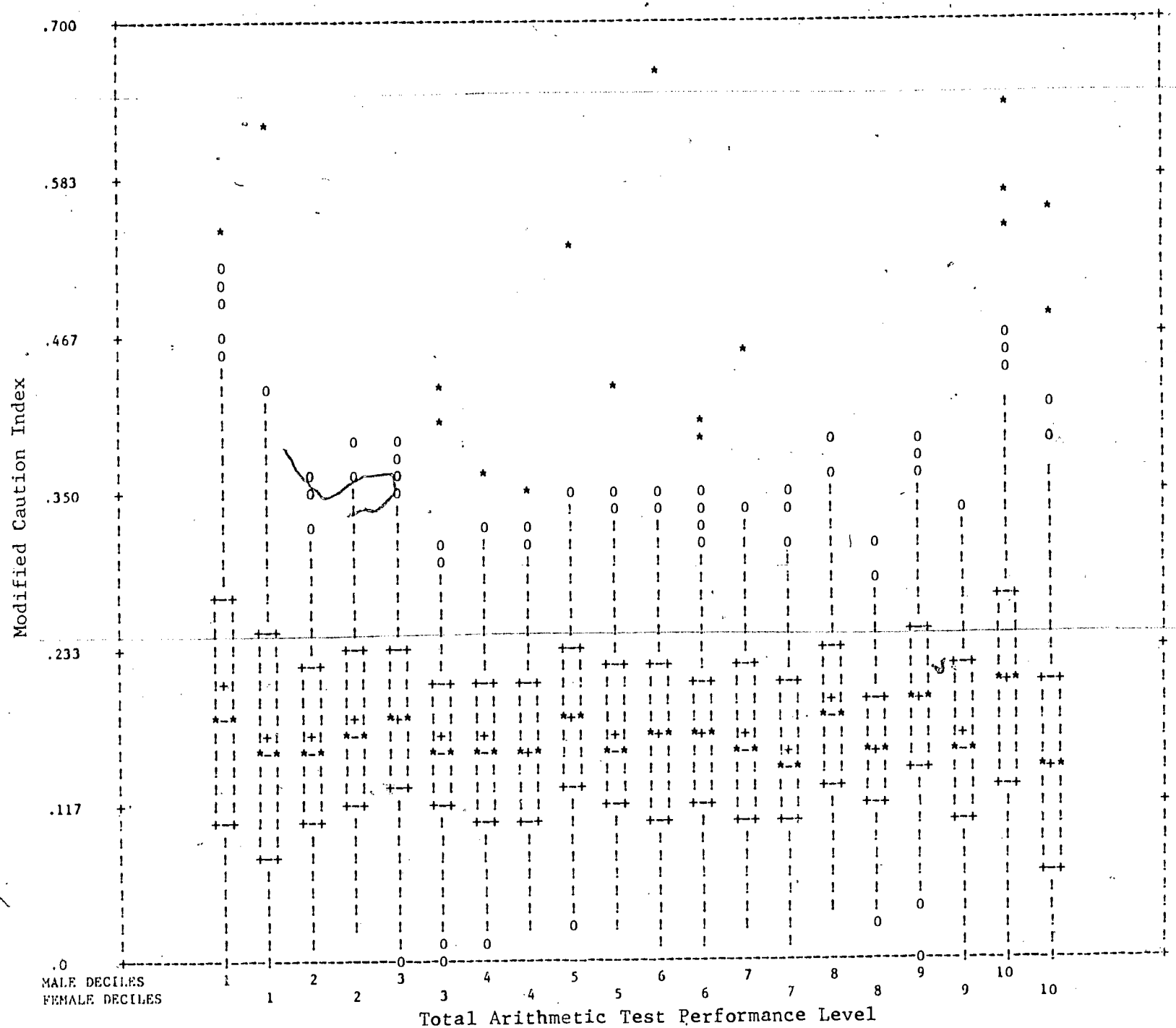

Figure 22. Box and whisker plot of modified caution index for males and females with the total arithmetic test score interval split into deciles. (Book 5, 13 year olds, 1978.)

APPENDICES

Appendix A

Classification, Percentage Correct, Sato's Caution Index, Modified Caution Index and the Point-Biserial for Cognitive Items

Book 4, 13 year olds, 1978
(N = 68)

Item Number	Item Classification*	Percentage Correct	Sato's Caution Index	Modified Caution Index	Point-Biserial
2	M	83.34	.45	.21	.34
3	A	71.69	.36	.17	.46
4A	A	50.96	.24	.12	.61
4B	A	63.97	.32	.15	.53
5A	A	97.17	.52	.24	.17
5B	A	94.54	.49	.23	.22
5C	A	84.16	.56	.26	.27
6A	G	84.08	.48	.22	.32
6B	G	23.43	.54	.28	.35
6C	G	34.39	.46	.24	.43
6D	G	19.45	.49	.26	.38
6E	G	45.84	.45	.23	.45
7	A	23.14	.42	.22	.44
8	M	39.68	.52	.26	.39
9	A	44.36	.42	.21	.47
10A	G	65.78	.77	.38	.18
10B	G	57.12	.65	.32	.28
10C	G	31.68	1.08	.56	-.06
10D	G	88.14	.55	.26	.25
10E	G	50.55	1.04	.52	-.03
11A	T	85.02	.55	.25	.27
11B	T	53.10	.34	.17	.53
11C	T	79.85	.28	.13	.48
12	A	24.33	.55	.29	.35
13	A	57.00	.79	.39	.17
14A	A	32.87	.46	.24	.43
14B	A	38.65	.42	.21	.47
14C	A	28.48	.42	.22	.45
14D	A	30.57	.40	.21	.48
15	A	64.87	.32	.15	.52
16	A	65.12	.35	.17	.50
17A	G	54.37	.41	.20	.47
17B	G	31.23	.54	.28	.37
17C	G	35.13	.63	.32	.30
18A	G	46.37	.70	.35	.24

Appendix A (continued)

Item Number	Item Classification*	Percentage Correct	Sato's Caution Index	Modified Caution Index	Point-Biserial
18B	G	17.28	.60	.31	.29
18C	G	16.50	.50	.27	.35
19	A	25.24	.58	.30	.32
20	A	86.83	.53	.25	.27
21A	A	7.35	.77	.40	.13
21B	A	24.29	.66	.35	.26
21C	A	19.45	.82	.44	.13
22	A	32.50	.31	.16	.55
23A	A	91.10	.36	.17	.33
23B	A	79.44	.46	.22	.36
23C	A	82.81	.43	.20	.36
24	G	29.13	.62	.32	.30
25A	A	68.24	.58	.28	.32
25B	A	25.77	.40	.21	.47
25C	A	28.44	.39	.20	.48
26	G	3.90	.22	.12	.35
27A	A	22.65	.40	.21	.46
27B	A	43.33	.44	.22	.45
28A	M	65.90	.49	.24	.39
28B	M	71.48	.51	.25	.36
29A	T	79.48	.31	.14	.46
29B	T	86.62	.27	.13	.43
29C	T	35.70	.68	.35	.26
30	A	16.78	.97	.52	.02
31	G	50.14	.59	.30	.33
32	A	33.28	.37	.19	.51
33	A	2.05	.31	.17	.24
34	G	47.97	.32	.16	.55
35	A	33.03	.42	.22	.46
36	T	55.31	.52	.26	.39
37	A	49.65	.37	.19	.51
38	A	26.14	.36	.19	.50
39	M	54.78	.47	.23	.42

*Item classification codes are A for arithmetic, G for geometry, T for tables and graphs, and M for miscellaneous.

Appendix B

Means, Standard Deviations, and Pearson Product Moment Correlations on Four Item Statistics

Book 4, 13 year olds, 1978

(N = 68)

Variable	Variable 1	2	3	4
1. Percentage Correct	1.00	.02	- .15	- .26
2. Point-Biserial		1.00	- .88	- .36
3. Sato's Caution Index			1.00	.99
4. Modified Caution Index				1.00
Mean	48.43	.36	.50	.25
Standard Deviation	25.08	.14	.18	.09

Appendix C

Classification, Percentage Correct, Sato's Caution Index, Modified Caution Index and the Point-Biserial for Cognitive Items

Book 5, 13 year olds, 1978

(N = 63)

Item Number	Item Classification*	Percentage Correct	Sato's Caution Index	Modified Caution Index	Point-Biserial
2	A	83.75	.52	.26	.32
3	A	43.30	.28	.14	.59
4A	A	94.80	.51	.27	.22
4B	A	78.59	.53	.27	.33
5A	G	88.18	.51	.26	.29
5B	G	82.37	.81	.41	.13
5C	G	76.44	.93	.46	.05
5D	G	87.29	.54	.27	.29
6A	M	57.07	.52	.26	.40
6B	M	53.33	.39	.20	.50
6C	M	47.97	.83	.41	.14
7A	T	69.70	.42	.21	.45
7B	T	66.45	.43	.21	.45
8A	A	38.91	.36	.18	.52
8B	A	35.95	.35	.17	.53
8C	A	42.40	.49	.25	.42
9	A	13.28	1.17	.58	-.10
10A	A	95.17	.41	.21	.26
10B	A	95.04	.41	.21	.26
10C	A	90.78	.41	.21	.33
11	T	42.40	.34	.17	.54
12	A	36.96	.30	.15	.57
13	A	22.22	.39	.19	.44
14	A	74.45	.43	.22	.42
15A	A	64.09	.31	.16	.55
15B	A	70.15	.34	.17	.51
15C	A	76.28	.27	.14	.54
16	A	19.21	.80	.40	.14
17	A	24.21	.29	.14	.52
18	A	58.69	.40	.20	.49
19A	T	40.78	.40	.20	.49
19B	T	51.06	.33	.16	.56
19C	T	42.65	.36	.18	.52
20A	M	1.62	.33	.16	.18
20B	M	2.88	.27	.13	.25

Appendix C (continued)

Item Number	Item Classification*	Percentage Correct	Sato's Caution Index	Modified Caution Index	Point-Biserial
21	G	82.66	.60	.30	.27
22	G	25.59	.40	.20	.45
23A	A	56.86	.57	.29	.35
23B	A	48.86	.49	.25	.42
23C	A	48.25	.43	.22	.47
24	G	62.27	.65	.33	.28
25	G	18.81	.83	.41	.11
26	A	54 47	.63	.32	.30
27	A	53.61	.39	.19	.51
28	M	8.90	1.08	.53	-.04
29	T	71.04	.45	.22	.43
30A	A	60.48	.16	.08	.68
30B	A	57.15	.18	.09	.67
30C	A	52.11	.17	.08	.69
30D	A	52.52	.23	.12	.64
30E	A	37.49	.22	.11	.63
31	G	9.46	.22	.11	.42
32	G	9.91	.63	.31	.21
33A	M	82.78	.58	.29	.29
33B	M	29.85	.90	.45	.08
34	A	71.12	.66	.33	.26
35	M	85.54	.28	.14	.46
36	A	84.24	.51	.26	.32
37	A	32.33	.34	.17	.52
38	A	81.48	.47	.24	.36
39	G	21.20	.45	.22	.39
40	A	45.94	.46	.23	.45
41	M	9.26	.59	.29	.22

*Item classification codes are A for arithmetic, G for Geometry, T for tables and graphs and M for miscellaneous.

Appendix D

Means, Standard Deviations, and Pearson Product Moment Correlations on Four Item Statistics

Book 5, 13 year olds, 1978

(N = 63)

Variable	Variable 1	2	3	4
1. Percentage Correct	1.00	.09	- .09	- .06
2. Point-Biserial		1.00	- .89	- .89
3. Sato's Caution Index			1.00	.99
4. Modified Caution Index				1.00
Mean	52.74	.38	.48	.24
Standard Deviation	26.13	.18	.21	.11