DOCUMENT RESUME

ED 068 498 TM 001 835

AUTHOR	Pike, Lewis W.; Evans, Franklin R.
TITLE	The Effects of Special Instruction for Three Kinds of Mathematics Aptitude Items.
INSTITUTION	Educational Testing Service, Princeton, N.J.
REPORT NO	CEEB-RB-72-19; CEEB-RDR-71-72-7
PUB DATE	May 72
NOTE	86p.
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	Admission Criteria; *Aptitude Tests; Grade 11; Instructional Design; Instructional Programs; *Instructional Systems; Mathematical Concepts; *Mathematics Instruction; Post Testing; Pretests; *Secondary School Mathematics; Sex Differences; *Statistical Analysis
IDENTIFIERS	Scholastic Aptitude Test

ABSTRACT

A differential instructional program was developed for each of three mathematics aptitude item formats to determine the relative susceptibility of each to special instruction. High school juniors were given a pretest that included items of each format and a parallel form as a posttest several weeks later. Between the pre- and posttests, experimental subjects received instruction in one of the three formats; controls received no special instruction. Pre- to posttest gains were analyzed using a two-way multivariate covariance analysis. The six dependent variables were the geometry and nongeometry posttest scores for each of the item formats. The seven covariates included the pretest scores corresponding to the six dependent variables and the Scholastic Aptitude Test-Verbal pretest score. Results of the statistical analysis showed that each of the three formats was susceptible to the special instruction directed toward it. The complex or novel item formats seemed to be more susceptible than the more straightforward item format. Female subjects tended to benefit less from the instruction than male volunteers. Mean gains of nearly one full standard deviation obtained by the "complex" format groups were considered to be of practical consequence and likely to influence admissions decisions. It is concluded that instruction was effective in two ways: subjects seemed to have learned a systematic approach to the item format and some basic mathematical concepts. (Author)

COLLEGE ENTRANCE EXAMINATION BOARD
RESEARCH AND DEVELOPMENT REPORTS

RDR-71-72, NO. 7

•

RESEARCH BULLETIN
RB-72-19 MAY 1972

The Effects of Special Instruction for Three Kinds of Mathematics Aptitude Items

•

Lewis W. Pike and Franklin R. Evans

EDUCATIONAL TESTING SERVICE
PRINCETON, NEW JERSEY
BERKELEY, CALIFORNIA

The Effects of Special Instruction for Three Kinds of Mathematics Aptitude Items

Lewis W. Pike and Franklin R. Evans

Educational Testing Service

Abstract

A different instructional program was developed for each of three mathematics aptitude item formats to determine the relative susceptibility of each to special instruction. Male and female high school junior volunteers in each of 12 schools were given a pretest composed of items of each format and a parallel form as a posttest several weeks later. In the intervening time experimental Ss received seven weeks (21 hours) of instruction directed at one of the three formats, while control Ss received no special instruction.

Pretest to posttest gains were analyzed in a two-way (sex by instructional group) multivariate analysis of covariance. The six dependent variables were the geometry and nongeometry posttest scores for each of the item formats. The seven covariates included the pretest scores corresponding to the six dependent variables and the SAT-verbal (SAT-V) pretest score.

Results of the statistical analysis showed that each of the three item formats was susceptible to the special instruction specifically directed toward it. The complex or novel item formats appeared to be more susceptible than the relatively straightforward item format. Female volunteers were found to be slightly less able mathematically at the outset and to benefit somewhat less from the instruction than male volunteers.

Mean gains of nearly a full standard deviation obtained by the groups instructed for the complex or novel formats were considered to be of practical consequence and likely to influence admission decisions.

The results of the study were consistent for all 12 schools. Although no group received instruction for the SAT-M per se, substantial pre- to posttest gains on that measure were also observed. Further analysis revealed that instruction, particularly for the complex or novel formats, was effective in at least two ways. Ss appeared to have learned a systematic approach to the item format as well as some very basic mathematical concepts.

The Effects of Special Instruction for Three Kinds of Mathematics Aptitude Items[1]

Lewis W. Pike and Franklin R. Evans[2]

Educational Testing Service

A new mathematics aptitude item format, Quantitative Comparisons (QC), has been considered as a possible replacement for the more traditional Regular Math (RM) and Data Sufficiency (DS) formats currently found in the mathematics section of the Scholastic Aptitude Test (SAT). A test made up of QC items that generally parallels the SAT-Mathematics (SAT-M) in mathematics content, item difficulties, and item _r_ biserials, has been found to require substantially less testing time per item, thus making it possible to obtain a given level of reliability in less time.[3] There has been some concern, however, that a test composed of QC items might be susceptible to the effects of practice, coaching, or other special instruction. The term, special instruction, will be used to refer to any program of instruction specifically designed to increase candidates' scores on an aptitude examination.

At the request of the College Entrance Examination Board, the present study was undertaken to determine the relative susceptibility of tests made up of QC, RM, or DS items to score changes resulting from special instruction. Information from this study may then be used as an aid in future planning for the SAT-M.

The QC, DS, and RM item formats are briefly described below. Examples and more detailed descriptions are provided later in the paper.

Each Quantitative Comparisons (QC) item presents the candidate with two quantities, one in column A and the other in column B, to be compared. The task is to make the comparisons and to mark A if the quantity in column A is the larger; B, if the quantity in column B is the larger; C, if the quantities are equal; or D, if there is not enough information to determine the quantitative relationship.

Each Data Sufficiency (DS) item presents the candidate with a question followed by two statements, labeled (1) and (2), in which certain data are given. The task is first to decide whether the question can be answered by A, (1) alone; B, (2) alone; C, (1) and (2) together; D, either (1) alone or (2) alone; or E, neither statement alone nor by (1) and (2) together; and then to mark the answer sheet accordingly.

Each Regular Mathematics (RM) item presents the candidate with a problem and five possible solutions. The task is to determine which of the five possible solutions is correct and to mark the answer sheet accordingly.

The General Problem of Special Instruction for Aptitude Tests

Although it has been found in several studies of special instruction for aptitude tests (College Entrance Examination Board, 1968; French, 1955; French & Dear, 1959) that expected score gains from such instruction are negligible, a widespread interest in special instruction and concern about its possible effectiveness persists. It is not surprising that this interest remains high, since standardized scholastic aptitude tests are widely used in making decisions of paramount importance to both students and colleges. Further, if special instruction substantially influences SAT scores, the validity of the test would be open to question, since it is intended to be a measure of relatively stable attributes developed over a long period of time (Carroll, 1970; College Entrance Examination Board, 1968; Whitla, 1962). The existence of an effective program of special instruction for an aptitude test would also suggest that, for many candidates, optimal preparation for that test should include some form of special instruction in addition to (but not instead of) the long-term development of the basic abilities being tested. Further, if such instruction were not equally available, there could be an unfair difference in how adequately students were prepared to take the test. The problem is compounded by the likelihood that the very students who cannot afford to attend a special coaching school are the ones least likely to attend schools where some form of special instruction for aptitude tests is provided as part of the curriculum.

Despite continued interest in the possible effectiveness of coaching for aptitude tests, one might see little reason for further research, given the generally negative conclusions of studies to date. The current question of the effects of special instruction on different mathematics aptitude item formats is not adequately answered by previous research, however, for several reasons. First, the majority of the studies were designed to determine whether coaching schools gave their clients an unfair advantage on the SAT. The instruction provided in these studies was, where one can ascertain its nature, rather scanty with little or no systematic attempt to identify the information and skills needed to perform well on the test and to develop materials to meet these needs. Since most previous research on this question involved Ss at the extremes of the ability range, generalization to the more heterogeneous population of candidates currently seeking admission to higher education becomes hazardous. Further, the DS format is rather complex, and as both Vernon (1954) and Loret (1960) have noted, complex formats appear to be more susceptible to practice and coaching than simpler ones. Similarly the QC format may be susceptible to practice and coaching because of its novelty as well as certain other characteristics peculiar to that format.

Studies of the Effects of Practice and Growth on SAT-M Scores

Before describing previous coaching studies in more detail it will be helpful to consider the effects of practice and growth on SAT-M scores. From studies of these questions, estimates may be obtained of the score gain that one could expect from merely retaking the test (practice) and from practice combined with the effect of additional schooling between tests (practice and growth).

Levine and Angoff (1958) observed an increment of about 10 SAT-M points (exclusive of growth) associated with a single practice session, and an additional 10 points following a second practice session. (The SAT-M scale has a range from 200 to 800, a mean at 500, and a standard deviation of 100.) No further gain was attributed to a third practice session and the results were similar for males and females. In a study of practice, growth, and coaching Frankel (1960) reported an effect for practice alone on SAT-M of about 29 points. He attributes the difference between his results and those of Levine and Angoff to equating error or to differences in the quality of the samples.

During the past several years, data for the combined effect of practice and growth have routinely been gathered for all candidates who repeat the SAT. Prior data on these combined effects (Frankel, 1960; Levine & Angoff, 1958; Watkins, 1958; Whitla, 1962) were all based on limited, specialized samples and will therefore be considered only when interpreting the findings of particular studies of special instruction.

Approximate mean changes in SAT-M scores for candidates taking the SAT in their junior year and again in their senior year are summarized over the academic years 1967-68 through 1970-71 in Table 1.[4] The over-

Insert Table 1 about here

all estimate of change is 15.1 SAT-M score points, based on over 1.6 million candidates, who had an average of five or six months of schooling between their first and second taking of the SAT.

Another recent statistical summary provided more detailed information on the combined effects of practice and growth by presenting score changes for all candidates repeating the SAT in the academic year 1969-70, according to their sex and their answers to two questions regarding their enrollment status in junior and senior mathematics courses.[5] The questions were: (a) Are you currently enrolled in a mathematics course?; and (b) Were you enrolled in a mathematics course last year (between September 1968 and June 1969)? The data are summarized in Table 2.

Insert Table 2 about here

As can be seen in Table 2, SAT-M score changes appear to be more related to the amount and recency of mathematics studies than to the candidates' sex. However, males showed a slightly larger increment than females as a result of practice and growth, even after mathematics enrollment was taken into account.

In summary, a single practice session has a small effect of about 10 points, and the combined effect of practice and of growth associated with five or six months of schooling between tests is about 15 points. The latter effect is slightly larger if the candidate is enrolled in mathematics courses in the junior and senior high school years, and tends to be slightly larger for males than for females.

Studies Related to Special Instruction for the SAT-M

In a review of the British literature on the effects of practice and coaching for intelligence tests (especially the "eleven-plus") Vernon (1954) concluded that: (a) a few hours of coaching and practice yield the maximum achievable gains; (b) more complex item formats are likely to be more coachable; (c) nonverbal tests are more likely to be coachable than verbal tests; and (d) coaching effects are greater for examinees who are less sophisticated about testing at the outset. He also pointed out that the general quality of instruction is likely to influence the outcome of a coaching experiment.

Jacobs (1966) conducted a study of the effectiveness of coaching for the College Board English Composition Test (ECT), an achievement test. It is of interest because of the relatively complex item types involved, because large gains were observed in some schools, and because of the differences in results from school to school which suggests a strong component of instruction. Student volunteers in six schools were randomly assigned to either a coached or a control group, while those in six other schools provided additional controls in case the coached students helped their control group classmates. The Ss' SAT-V and SAT-M scores each averaged about 500. Coaching was provided in the six schools in six three-hour sessions. About half the time was spent on instruction directed to the three ECT item formats, and the other half on English composition not directed specifically to the item formats. After the sixth week of instruction for experimental Ss, all Ss took an ECT. Ss in the six coached-and-control schools took a second ECT 10 months later. Jacobs found that coaching appeared to be effective initially in four of the six schools, with mean differences for coached Ss over control Ss ranging from 44 to 73 ECT scaled score points. (On the ECT scale, the mean is 500 and the SD is 100.) There was no appreciable difference in one school and a slight negative difference in the sixth school.

The study of coaching for the ECT offers insights into how an effective program might be provided for special SAT-M item types. First, the 18 hours of instruction was directed as much toward subject matter content as it was toward the ECT, per se. Second, the dramatic variations in coaching gains among the several schools (ranging from an apparent loss, to a 73 point gain), and in practice and growth during the following 10 months by the noncoached students (from mean gains of 15 points at one school to 94 at another), indicated that the specifics of the coaching and of the teaching situation may have a marked effect on the outcomes of a coaching experiment.

Several studies conducted between 1950 and 1965 directly investigated the effects of coaching on the SAT (Frankel, 1960; French, 1955; French &

Dear, 1959; Whitla, 1962). In general the investigators concluded that coaching did not result in substantial gains in SAT scores. A closer look at these studies, however, reveals several shortcomings which make generalization to the present study or to the current population seeking admission to higher education tenuous. First, all of these studies were conducted to determine if coaching as provided by commercial coaching schools or as part of a school's curriculum was effective in producing substantial gains in SAT scores. Since the present study was designed to determine if one item format is more susceptible to a thorough and systematic program of instruction than others, the results of these previous studies are not entirely applicable to the present question. Second, many of these studies were conducted during the 1950's when the SAT candidate population was much more homogeneous than it is today. Consequently, much of this research involved Ss enrolled in private preparatory schools or in public schools whose student bodies were well above the national average in tested ability. One can probably assume that the Ss in these studies were already quite sophisticated in standardized testing and were therefore less likely to benefit from special instruction than were students in the more general high school population. Third, the amount and quality of the instruction provided in these studies was, where one can ascertain its nature, rather scanty and unstructured. If one counts as much as two-thirds of the instructional time reported in these studies, as applying directly or indirectly to the SAT-M, then the estimated class time approached 20 hours in two of the studies and less than eight hours in each of the others. Although the instruction was probably adequate for testing the hypotheses under investigation at that time, one hesitates to rule out the possibility that a more thorough and systematic program of instruction could have produced substantial score gains. Finally, there are some rather puzzling contrasts among the results of these studies. For example, in two of these studies (Frankel, 1960; Whitla, 1962) the uncoached groups made substantial mean gains (up to 57 SAT-M points) between a pretest and a posttest. Such a finding, when contrasted with an expected gain from practice and growth of about 15 points, suggests that the schools these students attended were already doing exceptionally well in preparing their students for taking the test. Given that situation, it was not likely that a somewhat unstructured program of special instruction for the SAT would produce additional gains of any consequence.

A study by Roberts and Oppenheim (1966) represents a major shift in both the purpose of conducting a coaching study and the type of students studied. The study was undertaken to determine if coaching for the SAT could be effective in raising the SAT scores of students from academically disadvantaged backgrounds. The Ss were high school junior volunteers from 18 predominantly black high schools in Tennessee. Verbal instruction only was given in six schools, mathematics instruction only in eight schools, and no special instruction between pretest and posttest in the remaining four schools. Ss in each of the coached schools who indicated an interest in receiving the instruction were randomly assigned to either an instructed or a control group. The mean Preliminary Scholastic Aptitude Test-Mathematics (PSAT-M) pretest score of these Ss was approximately equivalent to an SAT-M score of 330. Instruction for the PSAT-M was provided in 15 half-hour sessions, using specially produced linear programmed materials under the direction of a classroom instructor. It had essentially no effect. The coached Ss gained the equivalent of approximately two SAT points, while the controls lost an average of about six points. In the discussion of their results, the authors pointed to the limited amount of instruction as a major limitation of the study. They also speculated that the control groups' loss of several SAT points was probably due to inadequate motivation.

Consideration for the Design of Special Instruction for the SAT-M

To complete the discussion of the general problem of special instruction for aptitude tests, it will be useful to consider some of the components contributing to a candidate's SAT-M score, and their implications for providing effective instruction. First is the candidate's underlying mathematics aptitude at the time he is tested. Presumably, instruction would not focus on this component, since it is assumed to reflect in large part abilities developed over a relatively long period of time. The second component is how well the candidate's mathematics skills cover the content domain to which the SAT-M is restricted. Thus, a part of effective instruction might well be to identify deficiencies through some diagnostic device and instruct for these specific gaps in the candidate's mathematics training that are likely to be tested on the SAT-M. Third is the degree to which the candidate is thoroughly familiar with appropriate topics. Without teaching him anything new about percentages, for example, proper instruction may help him to be "fresh" on the topic, and thus able to answer percentage questions more rapidly and accurately. Fourth is familiarity with that portion of the directions and procedures common to most objective tests. Structured practice with the test may be of most help here, but in at least two ways--using optimal guessing strategies and gaining a clear concept of the content sampled by the test--direct instruction could also be very helpful. Finally, there is the level of understanding of the specific item formats used in the test, particularly when some are rather complicated or novel. Effective instruction should help candidates to understand the tasks presented by the items, the meanings of the choice categories (where applicable), and the guessing strategies peculiar to given item types.

The specific problem to which the present study is addressed is the differential susceptibility of three types of mathematics aptitude test items to an intensive program of short term intervention. Of greatest interest is the susceptibility of Quantitative Comparison (QC) items to some form of special instruction. These items are more efficient than item types currently found in the SAT-M, but because of certain unique characteristics they represent a special kind of item that may be more susceptible to coaching than others. This presumed susceptibility has not yet been studied. Also of interest is the susceptibility of items in the Data Sufficiency (DS) format, which currently make up about 30% of the SAT-M. Because the DS format is complex, it might be expected to be susceptible to special instruction, but there is little evidence on that question. Of the coaching studies reviewed, only Whitla's and Roberts and Oppenheim's involved DS items, and these were not singled out for separate analyses. The Regular Math (RM) type of item would appear least likely to be susceptible to special instruction. It is included in the present study in part because doing so provides a link with past studies of coaching and primarily because the inclusion of all three item types will provide a solid basis for item-type selection for the SAT-M in the event that instruction is effective for one or more types.

Method

The selection of subjects and instructors as well as the design and development of instructional materials and curricula were consistent with the authors' intention of testing whether a program of short term instruction could, under ideal conditions, significantly increase mathematics aptitude test scores when tests are made up of QC, DS, or RM items. The strategy of providing ideal conditions was based on the observations that previous studies of coaching for score gains on aptitude tests have yielded consistently negative conclusions, and that there remains a general concern about whether special short term instruction can be effective.

Instructors and Schools

In order to ensure competent instruction, each of the 12 participating high schools was selected only after it was determined that there were two well qualified mathematics teachers on each faculty who were interested in the study and who had agreed to serve as instructors. The teachers were trained in a two-day workshop held shortly before instruction began. Five of the 24 teachers were women. All had some graduate training. The median of their teaching experience was approximately nine years.

Eight suburban and four urban public schools were included in the study. The suburban schools are all within a 50 mile radius of Princeton, New

Jersey. Two of the urban schools are in an Ohio city, and the remaining two in a city in western Pennsylvania.

Subjects

In September 1970, a general announcement was made in each participating school, asking eleventh graders to volunteer for special instruction directed toward the mathematics section of the SAT. The announcement specified that (a) not all who volunteered would be selected from the school; (b) most of those selected would be expected to attend seven instructional sessions, preceded by one testing session and followed by another, all to be held at the school on Saturday mornings; (c) the remainder of those selected would be expected to take part in the same testing sessions, but would receive instruction only after the second testing; and (d) all participating students would be allowed to take a regular SAT in April 1971 at no charge. The parents of each volunteer received an individual notice (see Appendix A) covering essentially the same points.

One instructor from each school completed a roster of all students who volunteered, giving their names and ninth and tenth grade mathematics courses and grades. In one urban school there were not enough volunteers, so additional subjects were recruited from another high school in the same city. In order to exclude volunteers who might be too sophisticated in mathematics to benefit from the special instruction, the names of students who were in an accelerated mathematics program or had an "A" average in mathematics were stricken from the rosters. Similarly, one student was dropped as unlikely to benefit from the instruction on the basis of failing grades in mathematics. After this initial screening, the list of volunteers for each school was sampled down to 54 by a random process to ensure that class size would not exceed 18.

The ratio of girls to boys among volunteers was about two to one. A similar ratio existed among those actually participating in the study.

Of the 555 Ss pretested, 509 remained in the project through the posttest. The dropout of less than 10% provided strong evidence of the Ss' motivation. The proportion of dropouts was nearly the same for each sex and for each of the four treatment groups. Their average pretest SAT-M score was 401, while that of the Ss who did not drop out was 408.

The Ss selected from each school were randomly assigned to two instructional groups and one control group, with the constraint that an equal number would be assigned to each group. The Ns and pretest SAT-V and SAT-M means for Ss in each group who continued through the posttest phase of the study are presented in Table 3.

Insert Table 3 about here

Procedure

For practical reasons, there were only two instructional groups in each school in addition to a control group. Because the question of the susceptibility of the QC format to special instruction was considered the most important, instruction for that item format was provided in all 12 schools (see Table 3). Instruction for DS items was given in six of the schools, and for RM items in the other six.

Within each school, each of the selected students was assigned at random to one of the S groups. The instructors notified the Ss of their group assignment by letter (see Appendix A). Changes from one group to another were not permitted. Students assigned to the control group were assured that instruction would be available to them after the posttest.

Pretests and posttests were administered at the participating high schools in October and December 1970, respectively, by personnel experienced in administering the regular SAT. The instructors were not present at these test sessions. For the pretest, administrators were sent rosters of students to be tested, with specific instructions for seating students and for distributing parallel Forms A and B alternately. By this procedure about half of the students received Form A and the other half Form B, in an approximately random assignment. The posttest was administered in such a manner that Ss took the alternate form of the one they had taken during the pretest. Such a design allows one to disregard the effects which could arise from the order in which test forms were administered. The assignment procedure was only partially followed in one school, resulting in seven Ss taking the same form at both administrations. The data for these Ss were excluded from the analyses.

Measures

Both Forms A and B of the pretest and posttest consisted of an SAT administered under standard conditions followed by a supplementary test made up of the new QC items. The QC test was constructed so that it was approximately parallel to the SAT-M in content as well as item difficulty and discrimination. Forms A and B each contained 55 QC, 43 DS, and 42 RM items.

In a factor analytic study of SAT-M content, Pruzek and Coffman (1966) found that most of the items that presented a geometric figure loaded heavily on the first factor, "Geometric Interpretation." Because of this finding, and the fact that about one-fourth to one-third of the items in each format present geometric figures, coaching effects were examined separately for geometry and nongeometry items within each item format.

Eleven pretest and 11 posttest scores were generated for each S, as listed below:

1. SAT-Verbal
2. SAT-Mathematics
3. QC geometry
4. QC nongeometry
5. DS geometry
6. DS nongeometry
7. RM geometry
8. RM nongeometry
9. QC total (3 + 4)
10. DS total (5 + 6)
11. RM total (7 + 8)

DS and RM items, which have five choices, were scored number right minus one-fourth the number wrong. QC items, which have four choices, were scored number right minus one-third the number wrong.

In addition to the scores described above, a third set of SAT-V and SAT-M scores was obtained for the 417 Ss who took the regular April 1971 SAT.

Instruction Common to All Three Curricula

Subjects in all three treatment groups were given 21 hours of instruction over seven Saturday mornings in November and December 1970. They were also given about 21 hours of homework, in the form of workbooks to be described below. Instruction for control Ss was provided after the posttest.

For descriptive purposes, the instruction common to all three curricula may be divided into two broad categories, test-taking skills and mathematics content. In fact, however, these components were integrated with one another in most of the instruction.

Test-taking skills were developed through test familiarization and the teaching of general test-taking strategies. The former included becoming thoroughly familiar with the general test directions and developing the ability to pace oneself realistically. The latter stressed an understanding of when to guess and how to make the best use of partial information.

Instruction in mathematics content was presented at three levels. At the first level, intensive drill was provided regarding the most basic mathematics facts, such as the squares of integers 1 through 13. In addition to ensuring a foundation for further instruction, this intensive drill was intended to reduce the S's error rate, while increasing his speed and confidence in answering mathematics questions. At the second level, instruction was provided to consolidate the S's knowledge and skill regarding a wider variety of tasks, such as computing percentages and solving simple equations. The third level involved "filling in the gaps" between the S's knowledge of mathematics and the content domain of the SAT-M. Some Ss, for example, had had virtually no instruction in problems dealing with inequalities, although these appear in all three item formats investigated.

It should be noted that very little of the content was completely new to most Ss. Thus most of the formal instruction was an intensive review, which was combined with an attempt to foster analytic skills, flexibility in mathematical thinking, and a systematic approach to problem solving that would make Ss more adept at answering mathematics test items.

Relatively uniform instruction was provided by the teachers, using a series of detailed lesson outlines for the appropriate item format, and other instructional materials prepared by the Educational Testing Service staff. A typical class session progressed through the following activities: (a) a review of the workbook lesson assigned as homework; (b) a 20-minute class test, which was immediately scored and discussed; (c) a "mini-lesson" presenting basic facts to be memorized; (d) instruction in selected mathematics content; (e) a brief diagnostic test of the content planned for the following lesson; and (f) an introduction to the workbook to be completed by the Ss as homework in preparation for the following class session. One or two "breaks" were provided at the teacher's discretion. A general introduction to the project was included in the first lesson. The final (seventh) lesson was devoted entirely to review of the previous six weeks' work.

The mathematics content presented through the mini-lessons and through the more general mathematics content instruction is outlined in Table 4.

Insert Table 4 about here

Note that geometry was covered in lessons I through IV in the mini-lessons, but reserved for lessons V and VI in the general content instruction. The basic reason for this was that some of the tests and workbook materials in the earlier lessons required elementary geometry. It also provided a change of pace within each three-hour class session.

Of the materials developed especially for the study, only the mini-lessons were entirely the same for all three curricula. Each mini-lesson was printed on 5" by 8" cards. At an appropriate time in each class session, the teacher reviewed previous mini-lessons and introduced the next one. Students were encouraged to keep their mini-lesson cards handy for ready reference and to learn thoroughly their contents. Mini-lessons 1 and 6 are shown in Appendix B. The information given in mini-lesson 1 is routinely provided as part of the directions for mathematics sections of the SAT.

The other instructional materials developed for the study included a separate set of workbooks, class tests, and lesson outlines for each curriculum.

The six workbooks developed for each curriculum constituted the backbone of the instructional materials. They were the basis for all homework except for study of the mini-lessons. The review of workbook assignments, and the discussion of questions arising from them, made up an important part of each class session. A new workbook was introduced and distributed at the end of each of the first six classes. Students were allowed to keep these for reference, until the seven weeks of instruction had ended.

The workbooks averaged about 75 pages (5 1/2" x 8 1/2") in length. The typical workbook was divided into four exercises, each consisting of problems or test items to be answered by the S. Answers and explanations for these questions were provided. Ss were encouraged to answer the questions first, then check their answers against those provided, and then to read over the explanations carefully. They were also asked to note any questions or explanations they did not understand, and to inquire about these in the next class session. The four exercises may be described as follows:

Exercise 1: This usually included 15 to 25 items of the appropriate format (QC, DS, or RM), all focused on the material learned in the previous class session. The answer and an explanation of each item was provided on the page immediately following that item.

Exercise 2: This consisted of problems related to mathematics content, such as understanding inequalities or knowing what information can be derived from figural problems. These problems were typically not put into the three item formats. Examples of these workbook materials are given in Appendix C.

Exercise 3: This consisted of items of the appropriate format, with the answer and explanation for each item given on the back of the page on which the item appeared. The content of these items covered material from all previous class sessions.

Exercise 4: For every workbook the last exercise was a practice test made up of items of the appropriate format. Item selection was based on the mathematics content and test-taking strategies taught up to that time. Standard directions and a separate answer sheet were provided. Ss were instructed to time themselves (20 minutes was usually the time allocated) and to complete the test with a different colored pencil after the time was up, if they were not already through. They were provided worksheets to compute two formula scores, one for the timed responses and the other for responses when there was no time limit. Answers and explanations were provided at the end of the practice test.

In most of the workbook exercises both content and strategies were presented through actual test items and through explanations of those items. Ss were required to make frequent responses to the material whether or not regular items were used, and immediate feedback beyond "right" or "wrong" was provided. By using test items and problems as a means of instruction, the workbooks also provided extensive practice in responding systematically and analytically to individual items under untimed conditions as well as practice in test-taking under timed conditions.

Instruction Unique to Each Curriculum

The RM, DS and QC curricula differed from one another primarily in that each provided instruction in test-taking strategies specifically applicable to its respective item format, and where needed, clarification of the directions associated with that item type. Most instruction was provided through test items, which were always of the format for which the curriculum was designed. Each item format and the specifics of instructions for it is described below.

Instruction for RM Items

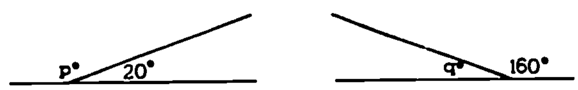
The RM format, which has been used continuously in the SAT since 1942 (Loret, 1960), is by far the most familiar of the three. It is seen in mathematics tests at virtually every educational level and is often used in teacher-made tests. The examinee's task is simply to solve the problem presented and to choose the correct answer from the five choices that are given.

Aside from the reference information given (see mini-lesson 1, in Appendix B) the directions for RM items are short and straightforward. They are given below, together with three examples of RM items.

Directions: **In this section solve each problem using any available space on the page for scratchwork. Then indicate the *one* correct answer in the appropriate space on the answer sheet.**

Example 1: **What is the weight of 28 feet of uniform wire if 154 feet weighs 11 pounds?**

(A) 2 lb. (B) $\frac{28}{11}$ lb. (C) $\frac{11}{2}$ lb. (D) 7 lb. (E) 14 lb.



Example 2:

In the figure, what is the sum p + q ?

(A) 20 (B) 140 (C) 160 (D) 180 (E) 540

Example 3: If n is a negative integer, which of the following numbers is the greatest?

(A) n^2 (B) $n - 1$ (C) $n + 1$ (D) $n^2 - 1$ (E) $n^2 - n$

Computation enters into answering most RM items. As in examples 1 and 2 above, many RM items present a problem to be solved, followed by five choices against which the solution can be checked. Example 3 is typical of other RM items in which a question is presented that refers directly to the five choices, often by an expression such as "which of the following . . .?"

Instruction for RM items stressed efficiency in problem solving. Skills in eliminating alternative choices were also emphasized. In explaining example 3, Ss can be shown that choices B and D can be readily eliminated, because it is apparent that B < C and that D < A. In general, Ss were shown that for many items some choices can be eliminated on such bases as whether they are odd or even, positive or negative, and so on.

An understanding of the effects of blind guessing and of informed guessing under conditions of formula scoring (the number right minus one-fourth the number wrong, for five choice items) was fostered by two classroom demonstrations. In the first, Ss blindly guessed the answers to test items when they had no information about each item. When their answers were tallied and formula scores applied, they saw that the average score for the class was around zero. In the second demonstration, they were asked to answer RM items when only partial information was available. For example, one item had as answers: (A) -4, (B) 0, (C) 1, (D) 4, and (E) 8. Each S was told that the answer was some positive number, and that he should choose one of the positive numbers. Through this demonstration Ss learned (often to their surprise) that choosing an answer when there was only partial information did indeed "pay off."

Instruction for DS Items

DS items have been included in the SAT since 1959. As the name Data Sufficiency implies, they focus on the S's ability to determine whether the data provided are sufficient to solve a problem, rather than on actually solving the problem through computation. The item format is complicated, and in that sense its introduction in the SAT runs counter to the general trend for item formats in the SAT to become less and less complicated (Loret, 1960). Loret noted that a major reason for reducing item complexity was to reduce the risk of coachability.

The full directions for DS items are given below. The actual operating directions are the five statements defining choices A through E.

Directions: Each of the data sufficiency problems below consists of a question and two statements, labeled (1) and (2), in which certain data are given. You have to decide whether the data given in the statements are *sufficient* for answering the question. Using the data given in the statements *plus* your knowledge of mathematics and everyday facts (such as the number of days in July or the meaning of *counterclockwise*), you are to blacken space

A if statement (1) ALONE is sufficient, but statement (2) alone is not sufficient to answer the question asked;

B if statement (2) ALONE is sufficient, but statement (1) alone is not sufficient to answer the question asked;

C if BOTH statements (1) and (2) TOGETHER are sufficient to answer the question asked, but NEITHER statement ALONE is sufficient;

D if EACH statement ALONE is sufficient to answer the question asked;

E if statements (1) and (2) TOGETHER are NOT sufficient to answer the question asked, and additional data specific to the problem are needed.

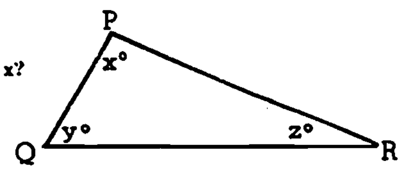
Note: A figure in a data sufficiency problem will conform to the information given in the question but will not necessarily conform to the additional information given in statements (1) and (2).

Example:

In ΔPQR, what is the value of x?

(1) PQ = PR

(2) y = 40



Explanation: According to statement (1), PQ = PR; therefore, ΔPQR is isosceles and y = z. Since $x + y + z = 180$, $x + 2y = 180$. Since statement (1) does not give a value for y, you cannot answer the question using statement (1) by itself. According to statement (2), y = 40; therefore, $x + z = 140$. Since statement (2) does not give a value for z, you cannot answer the question using statement (2) by itself. Using both statements together you can find y and z; therefore, you can find x, and the answer to the problem is (C).

The complexity of the DS format becomes apparent when one reads the directions. In their factor analysis of the SAT-M, Pruzek and Coffman (1966) commented that the Data Sufficiency factor ". . .appears to reflect the ability to read, interpret, and remember a rather extensive set of directions necessary for the solution of data sufficiency items [p. 10]."

Insert Figure 1 about here

The logic of the DS item is such that the choices are complexly interrelated, as shown schematically in Figure 1. Note, for example, that if statement

(1) alone is sufficient to answer the question, the correct choice must be A or D, so B, C, and E can be eliminated. On the other hand, if (1) alone is not sufficient, B, C, or E must be correct, and A and D can be eliminated. Another example of the interrelatedness of the choices arises when there is an item in which the required information is clearly divided between the two statements. In that case, neither statement alone is sufficient, and therefore the correct choice must be either C or E.

In addition to their greater complexity and the fact that the five choices are defined statements rather than answers to problems, DS items differ from RM items in the use made of geometric figures. For RM items, figures are drawn to scale unless it is otherwise indicated. For DS items, however, the directions specifically warn that the figures do not necessarily conform to the information given in the two statements.

The following rules were taught for answering DS items:

Rule I: It is seldom necessary to solve the problem completely.

Rule II: Always check both statements before answering.

Rule III: When you are not sure of the best choice, eliminate as many choices as you can and pick one of the remaining choices.

When answering an item such as the first DS example, many Ss tended first to solve for x before choosing an answer. Both in workbook and classroom exercises there were frequent reminders of steps to follow in order to avoid unnecessary computation as suggested in Rule I (see Appendix D). Another common failing was to choose A if statement (1) was sufficient, without considering statement (2), thus overlooking the possibility that the correct choice was D. Rule II and a general procedure for answering DS items that will be described below were directed to that problem. Much of the instruction was directed at specific ways to eliminate some of the choices for DS items, and Rule III was stressed in conjunction with that instruction.

The complexity of responding to DS items was substantially reduced by teaching a particular procedure for answering and by redefining the five choices in terms consistent with that procedure. The procedure was to write "YES," "NO," or "?" in front of each statement, indicating whether it alone was sufficient to answer the question, and then if both entries were "NO," to write under them "YES," "NO," or "?" for the sufficiency of statements (1) and (2) together. Then, the S was to choose an answer according to the pattern of the sufficiency statements he had just written. The DS choices were redefined in terms of the "YES," "NO," "?" pattern, as shown in Table 5. Ss were first drilled on the choice associated with each full-information pattern (1 to 5), then on the set of choices associated with each partial-information pattern (6 to 10).

Insert Table 5 about here

The effects of guessing on DS items under conditions of no information and of partial information were shown through a classroom demonstration described in Appendix E. All Ss in a class first completed a 20-item test with no information given, and then a similar test after having been provided with partial information such as that shown in the lower half of Table 5. Like the RM demonstration, the first part of the demonstration showed that the average score resulting from blind guessing was around zero. In the second part of the DS demonstration, the partial information that was provided would yield an expected score of about seven points, if it were used effectively. That the Ss had mastered the use of partial-information for effective guessing was indicated by the fact that an average gain of about six or seven points was observed for most of the DS classes when the demonstration was given.

Figures for DS items are usually generalized illustrations that are not drawn to scale. Therefore, Ss in the DS curriculum were taught which kinds of information could be assumed from such figures and which kinds could not. Such properties as linearity and betweeness could be assumed, for example, but the comparative magnitudes of the angles of a triangle could not.

Instruction for QC Items

The task presented by the QC item is to compare two quantities. The directions for the QC format, including two of the four examples provided, are as follows:

> ***Directions:*** **Each question in this section consists of two quantities, one in Column A and one in Column B. You are to compare the two quantities and on the answer sheet blacken space**
>
> **A if the quantity in Column A is the greater;**
> **B if the quantity in Column B is the greater;**
> **C if the two quantities are equal;**
> **D if the relationship cannot be determined from the information given.**
>
> ***Figures:*** **Position of points, angles, regions, and so forth can be assumed to be in the order shown.**
>
> **Figures are NOT NECESSARILY drawn to scale and may NOT agree to measures shown unless a note states that the figure is drawn to scale. Lines shown as straight can be assumed to be straight. Figures are assumed to lie in the plane unless otherwise indicated.**
>
> ***Note:*** **All numbers used are real numbers. In a question, information concerning one or both of the quantities to be compared is centered above the two columns. A symbol that appears in both columns represents the same thing in Column A as it does in Column B.**

	Column A		*Column B*	*Answers* A B C D
Example 1:	2×6		$2 + 6$	■ ▯ ▯ ▯
		$r + s = 1$ and $0 < r < 1$		A B C D
Example 2:	s		$1 - s$	▯ ▯ ▯ ■

QC items differ from both RM and DS items in having four rather than five choices. The basic operating directions for the QC format are the statements defining choices A through D. These are straightforward, although the information provided in the Figures and Note paragraphs may be somewhat complex.

The QC format is more like DS than RM in several interesting respects. First, the comparison of quantities can usually be carried out with a minimum of computation which is usually confined to simplifying complex expressions so that they are more readily comparable. Second, the fact that the choices are defined statements means that the RM procedure of solving and checking against the choices is not available. Third, figures accompanying QC items are typically not drawn to scale. Finally, the decision of whether to select choice D for a QC item is, in effect, to make a decision about "data sufficiency."

The QC format does not present some of the complexities found in DS items. There was, for example, no advantage in providing a redefinition of the choices, nor a need for stressing that it is unnecessary to compute a solution to a problem before selecting an answer. (A problem that appears to call for a solution is often part of the DS item, as shown in the first DS example, but this is not true for QC items.) With its focus on comparing quantities, however, the QC format lends itself to teaching several simplifying and time-saving techniques that apply to most QC items.

The first three rules taught to the QC Ss were as follows:

Rule I: When a problem involves only computation with numbers the answer is never D.

Rule II: When a problem involves unknowns, never forget about negatives and zero.

Rule III: Eliminate common terms from both sides.

A corollary to Rule I, of course, was to guess among the remaining three choices if the S̲ could not fully determine which one was correct. All three rules were frequently referred to in the explanations of QC items. An example of each rule is shown in the QC item explanations provided in Appendix F.

Because many of the QC items involved fractions, several procedures were taught to make these more readily comparable. One, of course, was to reduce one or both fractions to a simpler form. The opposite procedure was sometimes suggested for one of the fractions, so that either the numerators or the denominators of the two fractions were the same. Given the terms $\frac{24}{41}$ and $\frac{3}{5}$, for example, the right term can be "unreduced," so that the comparison is between $\frac{24}{41}$ and $\frac{24}{40}$. It is then evident that the right term is greater, making the answer **B.** As a final procedure, S̲s were shown that when both terms were positive fractions they could cross-multiply, so that the product of the extremes would become the left term, and the product of the means the right term. That is, the comparison of $\frac{a}{b}$ to $\frac{c}{d}$ is fully equivalent to that of ad to bc, if the original fractions are positive. For example, to compare $\frac{3}{5}$ and $\frac{4}{7}$, one can observe that $3 \times 7 > 5 \times 4$, so $\frac{3}{5} > \frac{4}{7}$.

In a review of guessing strategies similar to those in the RM and DS curricula, S̲s were reminded that whenever any choice could be eliminated, they should select one of the remaining choices rather than omit the item.

In addition to eliminating D by Rule I, they were shown that choice C can often be eliminated as well. For example, it may be obvious that $\frac{13}{119} \neq \frac{17}{155}$, but it is not immediately apparent that $\frac{13}{119} < \frac{17}{155}$.

As figures used with QC items are usually not drawn to scale, practice and instruction emphasized which information could be assumed from the figures provided, and which could not.

Analyses

This investigation focused on the relative susceptibility of the QC, DS, and RM mathematics aptitude item formats to special instruction. In the event that one or more of the item formats were found to be susceptible, it was also of interest to examine whether the susceptibility was related to the <u>S</u>'s sex. Therefore the data were analyzed in a two-factor multivariate analysis of covariance. The independent variables were instructional group (QC, DS, or RM instructed, and Control) and sex. The two independent variables, six dependent variables, and seven covariates for this multivariate analysis are listed in Table 6.

Insert Table 6 about here

Analysis was accomplished through the use of a multivariate analysis of variance computer program MANOVA described by Clyde, Cramer, and Sherin (1966). The program uses Wilk's lambda to test differences among the levels of a factor with respect to the set of dependent variables. In this case the analysis was performed on that part of the dependent variable set which could not be accounted for by the set of covariates. In this program the testing of effects proceeds from higher order interactions, through lower order interactions, to main effects within the hierarchical model. In addition to the multivariate tests of effects, the program provides a univariate analysis of variance for each dependent variable, discriminant function weights, discriminant scores, and correlations between each dependent variable and the discriminant scores, all of which can be used as an aid to the interpretation of results.

More subjective analyses were also conducted, in an effort to determine which aspects of the instruction appeared to be most effective. Item difficulties computed on the pretest data were compared to those computed on the posttest data, to identify items that were most affected by the instruction. Several of these items are presented and discussed.

Results and Discussion

Statistical analyses of the data pertaining to the research questions posed in the study were accomplished through multivariate analyses of covariance. The reader who is not familiar with multivariate analysis of variance and its interpretation is referred to Bock and Haggard (1968). To examine the practical significance of the effects of special instruction, pretest and posttest means and standard deviations were considered for various test scores. Finally, DS and QC items for which large changes in difficulty were observed were singled out for closer examination.

Statistical Significance

The program used for the analyses provides a test of whether the set of covariates is sufficiently related to the set of dependent variables to warrant the use of covariates. The first three canonical roots were significant. A battery reduction procedure (Hall, 1971) revealed that each covariate was related to the dependent variables in a unique way. Therefore, all seven variables were retained as covariates.

The statistical results for the treatment effects will be presented first, followed by those for the sex effect. The sex by treatment interaction will not be presented since the multivariate test of that interaction was not significant ($\underline{p} > .80$).

Main Effects Due to Treatments

The results of the multivariate analysis of covariance for the main effect due to treatment (instruction) are presented in Table 7. Each of the columns of the table presents the results for one of the three

Insert Table 7 about here

significant canonical variates. The first column results are for the canonical variate which provides maximum discrimination among the treatment groups, the second for the next most discriminating dimension and the third for the third most discriminating dimension. (Three discriminant functions is the maximum that could be obtained since there are only three degrees of freedom associated with the treatment factor.) Looking at the results this way allows us to assess the relative susceptibility of each of the three item formats, and for both geometry and nongeometry items within each format, to each of the three kinds of instruction. The effects of each treatment on the item format for which it was designed is of primary interest. The effects of each treatment on the other item formats and the relative effects of the treatments on geometry and nongeometry items are of secondary interest.

The multivariate F ratios presented in each column of Table 7 provide an exact test of differences among the different treatment groups with respect to the adjusted criterion variables along the optimal dimensions of discrimination. The values in the row labeled R_c^2 are the squared canonical correlations. These values are simply the squared product moment correlations between the maximally correlated linear combinations of the adjusted criterion variables and independent variables. The correlations (r_{cv}) between the adjusted criterion variables and the canonical variates shown in the middle portion of Table 7 indicate how much each dependent variable is contributing to the discrimination among the levels of the factor. A mean discriminant score for each group can be obtained by multiplying the discriminant function weights (not shown) and the adjusted criterion scores for each group. The mean discriminant scores for each group centered at zero, i.e., with the grand mean discriminant score removed, are shown in the lower portion of the table.

From the data in the first column of Table 7 it is possible to determine which of the three instructions was the most effective and for which dependent variables. Similarly, the data in the second column indicate the second most effective treatment, and so on. The QC subtests have the highest correlations with the first canonical variate (.82 for QC nongeometry and .52 for QC geometry). Therefore, it appears that the discrimination among the groups on the first canonical variate is due mostly to differences in scores on the two QC variables, and that differences on the QC geometry subtest are contributing somewhat less to the discrimination than are the QC nongeometry differences. As can be seen from the mean discriminant scores of the four groups on the first canonical variate, the QC instructed group received the highest mean discriminant score (.82) followed in order by the RM, DS, and Control groups.

The results in the second column indicate significant differences among the four groups on a dimension characterized mainly by differences in scores on the two DS subtests (the correlations between the second canonical variate and the DS nongeometry and geometry subtests are .81 and .49, respectively). Further, the DS instructed group earned the highest mean discriminant score (.58) on this dimension.

Similarly, the third canonical variate can be interpreted as an RM dimension. Unlike the QC and DS dimensions, scores on the RM geometry and nongeometry subtests correlate about equally with this dimension (.69 for RM geometry and .67 for RM nongeometry). As would be expected, the RM instructed group earned the highest mean discriminant score on this dimension.

Summarizing the information in Table 7 we can conclude that the instruction for each of the three item formats was effective. Subjects instructed for a given format gained more on a dimension characterized mainly by subtests composed exclusively of items of that format than either the control group or a group instructed for another format. We can also conclude that, due to the hierarchical nature of the analysis described above, the QC format is most affected by special instruction followed in order by the DS and RM formats. Geometry items in the QC and DS formats seem to account for somewhat less of the differences observed.

In addition to the analyses described above, three multivariate analyses of covariance, each representing a different set of orthogonal contrasts, were performed. In each of these analyses one instructed group was contrasted with the control group, while the other two instructed groups were contrasted with each other. The reader should keep in mind that the results of these three analyses are not independent of one another since the same groups were used in each analysis. Rather than present the full results of each of these analyses, the probabilities associated with the differences between each pair of groups have been summarized in Table 8.

Insert Table 8 about here

The probabilities associated with the multivariate F ratios and the six univariate F ratios for each special contrast are presented in Table 8. The multivariate p's simply indicate that there is a significant difference between the two groups being contrasted on the set of adjusted dependent variables. Of more interest, however, are the univariate p's. These indicate the dependent variables on which the contrasted groups differ. The underlying values in the upper half of Table 8 merely reflect the results of Table 7; i.e., instructed groups gained more than the control groups on items for which they were instructed. A question of more interest is: What is the relationship between two instructed groups with respect to the scores on the two subtests for which neither group was instructed? The underlined values in the lower half of Table 8 provide a direct answer to this question. That is, two instructed groups did not differ significantly with respect to subtests made up of the item format for which neither of those groups was instructed. In general, both groups showed some gains on these items but the gains were approximately equal for each group.

Main Effects Due to Sex

The results of the multivariate analysis of covariance for the main effects due to sex are presented in Table 9. The results are rather

Insert Table 9 about here

consistent, in that each of the six subtests appears to contribute toward the difference in the mean discriminant scores, in which males received a slightly higher score than females. That is, after the dependent variables are adjusted for differences in the covariates, males still obtain higher scores on each of the dependent variables. Note that the

DS subtests, as can be seen from the r_{cv} values in Table 9, contribute more than the other subtests to these differences. It should be pointed out that this analysis was directed primarily toward the gains from pre-test to posttest made by males and females on the various subtests. Inspection of a multivariate analysis of variance (not presented) of the unadjusted male-female criterion score differences did not indicate that some subtests contributed more to these differences than others. The females in the study performed at a slightly lower level on all of the posttests but when an adjustment was made for differences between the sexes in initial scores, the females appeared to perform at an even lower level than expected on the DS subtests.

The results for the sex effect presented in Table 9, and the absence of a significant sex by treatment interaction, suggest that all of the treatments were slightly more effective for males in the study than for females.

Practical Significance of Special Instruction

All of the foregoing results and discussion have focused on the statistical significance of the effects of special instruction on the geometry and nongeometry portions of the three subtests. Of equal importance is the practical significance of these effects. A common criterion for deciding whether coaching effects are of practical significance has been to compare the gains of the coached Ss to those of the uncoached Ss. Unless the gains of the coached Ss exceeded those of the uncoached Ss by at least one standard error of the test (about 30 points for the SAT-M), the coaching effect was considered to have no practical significance. This represents a difference of about 0.3 standard deviation units for the general SAT candidate population. In the present study a similar criterion may be used. However, since this study was designed to determine the relative susceptibility of the three item types to special instruction, the question of practical significance will be approached from the point of view of the gains to be expected if the SAT-M were composed entirely of a single item type. That is, each item type will be considered separately as if it were the only item type in the SAT-M.

The means and standard deviations of the pretest and posttest QC, DS, and RM total scores for each of the four treatment groups and for males and females separately within these four groups are presented in Table 10. Approximately half of the 502 Ss took Form A as a pretest

Insert Table 10 about here

then Form B as a posttest, while the remainder took Form B first then Form A. The means and standard deviations have been standardized to 50 and 10, respectively, in terms of the pretest data for each subset of each form. For example, the mean and standard deviation for each subtest (QC, DS, or RM) were computed for Form A for all Ss who took Form A as a pretest, regardless of sex or treatment group. These values were then used to compute conversion parameters for Form A, which were applied to the raw scores of each individual who took Form A as a pretest to obtain his pretest standard scores. These same parameters were applied to the posttest scores of those Ss who took Form A as a posttest to obtain their posttest standard scores. An identical procedure was carried out for Form B scores. The means and standard deviations presented in Table 10 are based on these "standard scores" combined for Forms A and B. Note that the posttest scores presented are in terms of the pretest "base" data for each group.

The sets of underlined means in each column of Table 10 provide an answer to the main question posed by this study, i.e., how much of a mean gain did each curriculum produce for a test in its respective item format, over and above the mean gain realized by a control group. Note that for QC items the mean gain for both males and females was nearly a full standard deviation, while the Control group gained only about one-quarter of a standard deviation. Similarly, males instructed for the DS items gained about one full standard deviation, but the uninstructed males gained only one-third of a standard deviation on the average. The mean gain for DS instructed females was somewhat less, about three-quarters of a standard deviation versus about one-third of a standard deviation for the uninstructed females, but the difference was still quite substantial. For both males and females, the mean gain scores of RM instructed groups were about one-half of a standard deviation compared to the uninstructed groups' mean gain of about one-fifth of a standard deviation.

Certainly the gains achieved by the QC and DS instructed groups on the appropriate items are of practical value. If the SAT-M were made up entirely of either of these item formats, the intensive program of instruction described earlier could be expected to produce changes in the scores that could result in different admissions decisions for many of the students. As expected, and as shown in the presentation of the statistical results, the gains achieved on RM items by the RM instructed group were smaller and thus of less practical consequence. It appears, then, that intensive special instruction as defined earlier in this paper produces both statistically and practically significant score gains on each of the three mathematics aptitude item formats. Also it appears that the QC and DS formats are more apt to be affected by such instruction than the relatively straightforward RM format. The latter finding, particularly with respect to the DS format, is consistent with Vernon's observation of the relationship between item complexity and the effects of practice and coaching on British intelligence tests.

Further inspection of the data in Table 10 provides additional insight into the sex factor noted earlier, and the relationship among the gains achieved by instructed groups on tests composed of items for which the groups were not instructed. In general the females were somewhat less able than males in mathematics at the outset, and they did not appear to gain as much from the instruction.

There does not appear to be any greater difference in the gain scores of males and females for any one treatment, which confirms the absence of a treatment by sex interaction. However, females did gain somewhat less than males on DS items regardless of the instructional group of which they were members. The results in Table 10 also confirm the observation that two instructed groups did not differ substantially from one another with respect to gains on tests composed of items for which neither group was coached. For example, both RM and DS instructed Ss made about one-half of a standard deviation gain on QC items.

Gulliksen (1950), in a discussion of intrinsic validity, describes coaching as harmful if the coaching changes the test score without a resultant change in the Ss' ability to perform on the criterion that the test was designed to predict. Although our data do not provide a direct answer to this question, there is reason to believe that the instruction provided all Ss in this study probably did equip them to perform better in mathematics courses at a higher level. Throughout the instruction emphasis was given to the basic mathematical concepts upon which all higher mathematics are based. This strategy was a deliberate attempt to provide a strong test of whether a program of special instruction would be effective in producing score changes. The results of this research then show only that the program of special instruction as defined was effective. Therefore, the question of whether such a program had either a detrimental or a positive effect on the validity of the test remains unanswered. Although strong statements about the effects of coaching have been made based on the studies cited earlier in this paper, those studies were also inadequately designed to answer the important question of the effects of coaching on the validity of the test.

Stability of Results Across Schools

One may reasonably ask whether the gains in test scores were the result of exceptionally good teaching in a relatively small subset of the participating schools. Inspection of the school by school data (not presented) revealed that in every school the two instructed groups gained more on the items for which they were instructed than did the Control group. In general the QC instructed group in each school made larger gains on QC items than the RM or DS instructed group in the same school made on the items for which they were instructed.

Changes in SAT-V and SAT-M Scores

Although this study was not designed specifically to investigate increases in SAT scores per se, or the long term effects of instruction on these scores, it does provide some data bearing on these questions. The reader should keep in mind that no group received instruction for the full SAT-M, but rather for one mathematics aptitude item format plus the mathematics content considered necessary to perform adequately on tests of mathematics aptitude. Of the 502 Ss who took the special pretest and posttest SATs, 417 also took the SAT a third time in the regular April 1971 administration. The SAT-V and SAT-M means and standard deviations for these Ss are presented in Table 11.

Insert Table 11 about here

Although this group of Ss is a self-selected sample of the total study sample, the reader will note from the pretest SAT-V and SAT-M means and standard deviations that the four groups did not differ substantially from those presented for the total sample in Table 3. Therefore, the data are probably representative of the total study sample. There are several interesting aspects of Table 11. First, note the relative magnitude of the SAT-M changes from October to December. The RM instructed group gained most during that period, followed by the DS, QC, and Control groups, in that order. This is exactly the expected order since the SAT-M contained twice as many RM as DS items, and no QC items. Further, the Control group received no special instruction during that period. Thus, one might conclude that had instruction been given for the SAT-M itself, the gains would have been of practical value. Second, the data in the October to April columns reflect the combined effects of instruction, two practice sessions and about five months of additional schooling for all four groups. During the period January through March the Control group received instruction for the RM format, the DS format, or some combination of these two. The curriculum for the Control groups was probably more closely related to the SAT-M than were any of the other three curricula. The October to April SAT-M gain averaged across all four groups was approximately 50 points. A best estimate of the gain due to two practice sessions and growth is in the neighborhood of 30 points (see the discussion on the effects of practice and growth earlier in this paper). Based on that estimate, about 20 of the 50 points observed here could be attributed to the special instruction. This figure is still substantial considering that three (RM, DS, and QC) of the four groups received instruction not specifically related to the SAT-M.

It appears from data provided by Donlon and Angoff (1971), Levine and Angoff (1958), and Watkins (1958) that the approximately 33 point gain in SAT-V score from the October to the April testing is mostly the result of the two practice sessions plus an additional five months of schooling. Therefore, it does not appear that the mathematics instruction had any substantial effect on the SAT-V scores of the Ss in this study.

Effects of Instruction on Individual Test Items

The analyses presented below were performed to gain insights into the mechanisms by which the various aspects of the curricula were effective. To examine the effects of special instruction at the item level, pretest and posttest item difficulties (deltas)[6] were computed for each item. For any given item, the deltas were based on Ss instructed for its item format. For each QC item, for example, the pretest delta was computed on QC instructed Ss who had received that item in the October pretest, and the posttest delta was computed on QC Ss who received it on the December posttest. The resulting values were plotted for the QC, DS, and RM item formats and are presented in Figures 2a, 2b, and 2c respectively. The

Insert Figures 2a, 2b, and 2c about here

solid line on the 45° diagonal represents no change. Points that fall above that line represent items which relatively fewer Ss answered correctly following instruction, while points below the line represent items which relatively more Ss answered correctly following instruction. (Three QC, one DS, and four RM items are not shown because either too few Ss responded or too few Ss responded correctly to allow the computation of the pretest delta.) The broken lines above and below the solid diagonal represent an increase or decrease in item difficulty of two deltas. As the standard deviation of delta is four, items falling outside the band formed by the broken lines are those which changed in difficulty by more than one half of a standard deviation.

In examining the scatterplots note that the points on all three tend to cluster somewhat below the line of no change, indicating that items on all three subtests tended to be easier after the Ss received appropriate instruction. This downward trend is more pronounced for the QC and DS items than for the more straightforward RM items, reflecting the previously noted result that the more complex item formats are most susceptible to special instruction. It is also interesting to note that the instruction was successful over a wide range of item difficulties. For example, there are instances in every plot where items with a pretest delta as low as 8 or 9 or as high as 17 or 18 became easier after appropriate instruction.

In Figures 2a and 2b, and to a lesser extent 2c, there are a substantial number of items represented by points which fell well below the lower dashed line. Several of these items were inspected in order to gain some insight into why they became so much easier after instruction. Such insight will be valuable to those who construct the test, and should help determine how various aspects of the instruction were operating.

Effects of Instruction on Selected QC Items

Examples of QC items which dropped considerably in difficulty are presented in Table 12, together with pretest and posttest item deltas. The pretest and posttest response patterns are also presented for each item, with the frequencies of the correct choices underlined.

Insert Table 12 about here

In examining the response frequencies, note that five of the eight examples have D as the correct response. D-keyed items probably represent, to the uninitiated, the most difficult of all of the QC items, because respondents are probably not familiar with deciding whether there is enough information to determine the relative magnitude of two quantities. In most mathematics curricula students are taught to determine the value of a quantity. When confronted with an item for which they are unable to determine a value they hesitate to decide that D, "the relationship cannot be determined," is the correct choice. Faced with this situation it appears that Ss tended to use partial information to try to decide that one quantity was larger than the other or that the quantities were equal.

A good way to interpret the data in Table 12 is to look at the most popular incorrect response before instruction and try to determine the reason for its popularity. On D-keyed items II, III, IV, and VII, for example, there was a definite tendency to choose A or B. The basis for that tendency seems to have been intuitive rather than rational, as if the incorrect responses had been made on the basis of some kind of vague feeling based on misapplied partial information. In item II, for example, the Column B quantity may have seemed larger to many pretest Ss because it has the most terms. The Column A value in item III may have appeared larger for the same reason. In item IV, Column A may have appeared greater because the domain of x extends to plus one, while that of y in Column B extends only to zero, and in item VII, the Column A value may have seemed larger because it appears to be positive, while the Column B value appears to be negative.

How did the instruction correct these inappropriate responses? The answer is probably twofold. First, Ss were taught to respond to the items analytically and systematically, rather than on some overgeneralized and incomplete basis. Second, they were taught the basic mathematics content necessary to respond systematically to the items. On item II, for example, most posttest Ss would probably first simplify the Column B expression to unity, so the comparison is then between $\frac{x}{y}$ and 1. They would then note that since x and y are unknowns, the relationship between $\frac{x}{y}$ and 1 cannot be determined. Similarly, posttest Ss would tend to simplify the terms in item III to $\frac{1}{xyz}$ and $\frac{xyz}{3}$, and again recognize that since unknowns remain, the relationship between the two quantities cannot be determined. In responding to item VII, instructed Ss probably tended to apply a particular strategy they were taught; to consider positive, zero, and negative values when comparing unknowns. For example, if x and y were positive, $x + y$ would be greater than $-x - y$, but if the unknowns were negative, the relationship would be reversed. Therefore, the correct answer to item VII would have to be D.

Item I provides an interesting example of rather straightforward items on which pretest Ss seem to have faltered for the lack of a systematic approach. Choice C probably attracted many who thought that since the numbers in both columns were the same the terms must be equal, but who failed to recognize that although the sums were the same, the products were not. Posttest Ss would be more likely to simplify both expressions first by eliminating the common value (268), so that the terms to be compared became 3 X 8 and 2 X 9. At that point, it becomes obvious that A must be the correct choice.

Thoroughly learning some rather simple content probably made the most difference in items V and VIII. In item V, learning and applying the fact that the interior angles of a triangle must sum to 180° was probably the key to the large change in item difficulty. For item VIII, the needed information was the fact that a real number may have either a positive or a negative root.

It appears, then, that the instruction was effective for these items in several ways. First, Ss were taught to approach the problem systematically, to make their comparison task simpler by stripping away the unnecessary clutter. Second, they were taught the necessary basic facts and how to use these in the QC task of comparing two quantities. Instruction dealing with exponents, basic geometry, fractions, and the evaluation of inequalities and unknown quantities was probably particularly helpful to Ss both in simplifying expressions and in comparing quantities (or recognizing their noncomparability).

Effects of Instruction on Selected DS Items

Examples of DS items which instructed Ss answered much more successfully after than before instruction are presented in Table 13, together with pretest and posttest item deltas. Pre- and posttest response patterns

Insert Table 13 about here

for each item are also given, again with the frequencies of the correct choices underlined. The choices for the DS items are defined as follows:

A. if statement (1) ALONE is sufficient, but statement (2) alone is not sufficient to answer the question asked;

B. if statement (2) ALONE is sufficient, but statement (1) alone is not sufficient to answer the question asked;

C. if BOTH statements (1) and (2) TOGETHER are sufficient to answer the question asked, but NEITHER statement ALONE is sufficient;

D. if EACH statement ALONE is sufficient to answer the question asked;

E. if statements (1) and (2) TOGETHER are NOT sufficient to answer the question asked, and additional data specific to the problem are needed.

The main strategy Ss were taught for responding to DS items was to use partial information to eliminate some choices and then to guess among the remaining choices. As shown earlier (see Table 5), the number of possible choices is reduced to 2 or 3, whenever one of the sufficiency statements can be answered but not the other. When it can be determined that neither statement alone is sufficient, but not whether the two statements together are sufficient, the number of possible choices is reduced to 2. If the strategy of using such partial information was well learned, posttest item data should show a trend toward fewer Ss omitting an item, and toward specific patterns of incorrect response. For the items included in Table 13, the rate of omitting dropped by about two-thirds, from pretest to posttest. The drop was most noticeable for items III and VIII. The trend toward specific patterns of error response was even more pronounced. All of the items in Table 13 are keyed either C or E. If the appropriate guessing strategy is used, the most popular wrong response for C-keyed items would be E, since Ss having partial information on those items would be guessing within choice patterns (C or E), (A, C, or E), or (B, C, or E), depending on the partial information used. Similarly, the most popular wrong response for E-keyed items should be C. Although there is some evidence of a predisposition to use this strategy before instruction, the pattern is much stronger for the posttest data. In fact, for items I, III, V, VI, and VIII, virtually everyone answered either C or E on the posttest.

Observation of DS classes early in the instruction period revealed that prior to instruction Ss had a tendency to use information from one sufficiency statement when deciding if the other statement provided sufficient information to solve the problem. Item I provides a rather good example of the results of such a faulty strategy. Note that 7 of the 25 Ss who answered item I incorrectly on the pretest chose A. These Ss probably reasoned that there was only one even number, namely 6, whose square was between 20 and 50. Thus it appears that they used the fact that the number is even, given in statement (2), when evaluating statement (1). Note that after instruction only one S chose A. For this rather easy item, at least, the warning to consider each statement separately was apparently effective. Item VIII provides another example of the same faulty strategy. Apparently many Ss used the information from statement (1) that $\triangle$OPQ was equilateral in arriving at B as their answer. Again, only one S made such a choice after the instruction.

Although the test directions specifically warned Ss against making assumptions about the figures used in DS geometry items, there was a definite tendency for them to do so prior to the special instruction. In item II, for example, it appears from the figure that lines $\overline{AB}$ and $\overline{DC}$ are parallel. If one makes that assumption, then uses the fact that angles q and r would be alternate interior angles, he may conclude that B is the correct choice. Eleven Ss apparently made such an assumption for item II. Although 7 Ss still chose B after instruction, the trend was in the expected direction. It is also interesting to note that after instruction more Ss were able to distinguish correctly between C and E as the correct choice for item II, suggesting that the geometry content instruction was effective.

Item V provides an example of an item where Ss apparently did not fully understand their task. The tendency for incorrect responders to choose A prior to instruction could have been due to their thinking in terms of gross income instead of net profit. In that case, of course, statement (1) provides sufficient information and A is a logical choice. One aspect of the DS instruction was to help the S to determine what it was he was being asked.

It appears that the DS instruction, like that for QC, was effective because of a combination of item specific strategies and basic mathematics content. Instruction aimed at helping an S cope with the complexity of the format and use what partial information he had to make a reasoned response was particularly helpful. Once he understood how to approach the item, the content instruction enabled him to gain more partial information from the data presented and thus, in the case where he could not fully determine the answer, to respond on an intelligent basis.

Summary and Conclusions

A different instructional program was developed for each of three mathematics aptitude item formats to determine the relative susceptibility of each to special instruction. Male and female high school junior volunteers in each of 12 schools were given a pretest composed of items of each format and a parallel form as a posttest several weeks later. In the intervening time experimental Ss received seven weeks (21 hours) of instruction directed at one of the three formats while control Ss received no special instruction.

The statistical significance of the gains from pretest to posttest were analyzed in a two-way (sex by instructional group) multivariate analysis of covariance. The six dependent variables were the geometry and nongeometry posttest scores for each of the item formats. The seven covariates included the pretest scores corresponding to the six dependent variables, and the SAT-V pretest score.

Results of the statistical analysis showed that each of the three item formats was susceptible to the special instruction specifically directed toward it. The complex or novel item formats appeared to be more susceptible than the relatively straightforward item format. Female volunteers were found to be slightly less able mathematically at the outset and to benefit somewhat less from the instruction than male volunteers. Mean gains of nearly a full standard deviation obtained by the groups instructed for the complex or novel formats were considered to be of practical consequence and likely to influence admission decisions.

The results of the study were consistent for all 12 schools. Although no group received instruction for the SAT-M per se, substantial pre- to posttest gains on that measure were also observed. Further analysis revealed that instruction, particularly for the complex or novel formats, was effective in at least two ways. Ss appeared to have learned a systematic approach to the item format as well as some very basic mathematical concepts.

References

Bock, R. D. & Haggard, E. A. The use of multivariate analysis of variance in behavioral research. In D. K. Whitla (Ed.), Handbook of measurement and assessment in behavioral sciences. Reading, Mass.: Addison-Wesley, 1968. Pp. 100-142.

Carroll, J. B. Possible directions in which College Board tests of abilities and learning capacities might be developed. In College Entrance Examination Board, Report of the commission on tests: II Briefs. New York, N. Y.: College Board, 1970.

Clyde, D. J., Cramer, E. M., & Sherin, R. J. Multivariate statistical programs. Coral Gables, Fla.: Biometric Laboratory of the University of Miami, 1966.

College Entrance Examination Board. Effects of coaching on Scholastic Aptitude Test scores: Statement by College Board Trustees. New York, N. Y.: College Board, 1965, 1968. Pp. 8-11.

Donlon, T. F. & Angoff, W. H. The Scholastic Aptitude Test. In W. H. Angoff (Ed.), The College Board Admissions Testing Program. New York, N. Y.: College Entrance Examination Board, 1971. Pp. 15-47.

Frankel, E. Effects of growth, practice, and coaching on Scholastic Aptitude Test scores. Personnel and Guidance Journal, 1960, 38, 713-719.

French, J. W. An answer to test coaching. College Board Review, 1955, 27, 5-7.

French, J. W. & Dear, R. E. Effect of coaching on an aptitude test. Educational and Psychological Measurement, 1959, 29, 319-330.

Gulliksen, H. Intrinsic validity. American Psychologist, 1950, 5, 511-517.

Hall, C. E. Generalizing the Wherry-Doolittle battery reduction procedure to canonical correlation. Journal of Experimental Education, 1971, 39, 47-51.

Jacobs, P. I. Effects of coaching on the College Board English Composition Test. Educational and Psychological Measurement, 1966, 26, 55-67.

Levine, R. A. & Angoff, W. H. The effects of practice and growth on scores on the Scholastic Aptitude Test. Statistical Report 58-6. Princeton, N. J.: Educational Testing Service, 1958.

Loret, P. G. A history of the content of the Scholastic Aptitude Test. College Entrance Examination Board, Research and Development Reports and ETS Test and Development Memorandum 60-1. Princeton, N. J.: Educational Testing Service, 1960.

Pruzek, R. M. & Coffman, W. E. A factor analysis of the mathematical sections of the Scholastic Aptitude Test. College Entrance Examination Board, Research and Development Report 65-6, No. 10 and ETS Research Bulletin 66-12. Princeton, N. J.: Educational Testing Service, 1966.

Roberts, S. O. & Oppenheim, D. B. The effect of special instruction upon test performance of high school students in Tennessee. Research Bulletin 66-36. Princeton, N. J.: Educational Testing Service, 1966.

Vernon, P. Symposium on the effects of coaching and practice in intelligence tests. British Journal of Educational Psychology, 1954, 24, 57-63.

Watkins, R. W. Change in Scholastic Aptitude Test scores of candidates tested in the junior and senior years. Statistical Report 58-7. Princeton, N. J.: Educational Testing Service, 1958.

Whitla, D. K. Effect of tutoring on Scholastic Aptitude Test scores. *Personnel and Guidance Journal*, 1962, *41*, 32-37.

Footnotes

[1] This research was funded by the College Entrance Examination Board.

[2] The authors wish to express their gratitude for the support provided by staff members of the Developmental Research Division and the Mathematics Department of the Test Development Division of Educational Testing Service, to S. K. Damarin, who contributed significantly to the design of the instructional materials and curricula, to F. R. Creech and C. E. Hall, who provided valuable assistance in data analysis and interpretation of results and to the teachers in the participating schools.

[3] A special study on these questions was conducted by the Test Development Division of Educational Testing Service and reported in an interoffice memorandum from J. S. Braswell dated 4/6/71.

[4] The data presented in Table 1 were adapted from the routine 5 year score change summaries for 1967 to 1971 and were reported in memoranda from A. L. Hussein and J. Stern dated 1/17/72, 2/9/72, and 2/16/72.

[5] The data presented in Table 2 were adapted from a special study of SAT repeaters reported in an Educational Testing Service interoffice memorandum from A. L. Hussein dated 1/12/71.

[6] The "delta" scale of item difficulty is a normalized transformation of the more usual "percent-pass" index, with a mean of 13 and a standard deviation of 4. High values on the delta scale indicate a high level of difficulty rather than the reverse, which is true of the percent-pass scale. For a further explanation of the delta

index, see: Conrad, H. S. _Characteristics and uses of item-analysis data_. _Psychological Monograph_, No. 295. Washington: American Psychological Association, 1948.

Table 1

Approximate Mean Changes in SAT-M Scores for Candidates Taking the SAT in Their Junior Year and Again in Their Senior Year, during the Years 1967 to 1971

Senior year administration	Junior year administration							
	March		April or May		July		Totals	
	N[a]	Mean change	N	Mean change	N	Mean change	N	Mean change
November	166	16.4	476	11.6	94	13.7	737	13.0
December	179	20.9	567	15.6	86	18.6	833	17.1
January	17	19.1	60	14.2	14	16.0	90	15.4
Totals	362	18.7	1,102	13.7	195	16.0	1,660	15.1

[a] In thousands

Table 2

Mean SAT-M Score Changes from Junior to Senior Year for Candidates According to Sex and Current and Past Mathematics Enrollment Status[a]

Enrollment status		Mean change		
Currently taking mathematics?	Mathematics courses last year?	Male (N=191,000)	Female (N=173,000)	Total (N=364,000)
yes	yes	22.6	20.2	21.7
yes	no	17.9	17.2	17.5
no	yes	10.9	10.2	10.5
no	no	7.1	4.5	5.2
	Total	20.1	15.3	17.8

[a] Data are based on all candidates who repeated the SAT in November or December 1969, or January 1970.

Table 3

Mean Pretest SAT-V and SAT-M Scores for the Instructed and Control Groups in the Participating Schools

School	Subject group											
	QC instructed			DS instructed			RM instructed			Control		
		Mean score			Mean score			Mean score			Mean score	
	N	SAT-V	SAT-M	N	SAT-V	SAT-M	N	SAT-V	SAT-M	N	SAT-V	SAT-M
1	15	382	423				15	403	410	13	408	403
2	14	401	448				13	415	465	13	420	420
3	14	376	440				12	370	429	14	384	430
4	15	437	393	12	359	352				14	405	384
5[a]	12	382	395	16	400	410				14	401	409
6	15	429	408	16	397	417				14	378	381
7	16	366	354				16	370	381	15	352	355
8	16	433	447				14	417	454	14	391	420
9	11	415	428				9	371	390	10	402	429
10	12	478	460	12	398	434				14	382	402
11	14	368	381	15	397	378				14	377	384
12	16	374	411	17	396	396				16	430	398
Total	174	403	414	88	392	398	79	392	422	168	394	400

[a] Data for 4 QC and 3 Control Ss in school 5 were excluded because they were given the same test form for both pretest and posttest.

Table 4

Mathematics Content Outline

Lesson	General content	Mini-lesson
I	Terminology and symbols, especially inequalities. Simple linear equations of one variable. Substitution.	Formulas for the area and perimeter of triangles and circles. Pythagorean theorem. Definitions of symbols.
II	Properties of roots and powers. Mixed fundamental operations. Simple properties of fractions.	Vertical angles and supplementary angles.
III	Whole number properties: odd, even, consecutive, divisibility, prime, squares of integers. Averages.	Parallel lines intersected by a transverse.
IV	Fractions, as ratios and percentages. Proportions. Linear equations. Difference of squares.	Special right triangles: angles of 30-60-90, sides of 3-4-5, *et cetera*.
V	Plane geometry: lines, angles, triangles.	Review of whole number properties: odds and evens, prime numbers, perfect squares.
VI	Plane geometry: circles, polygons, miscellaneous.	Whole number properties: consecutive integers, averages.
VII	General review.	General review.

Table 5

DS Choices Associated with Patterns of Complete and Partial Information

Pattern	Is statement (1) ALONE enough?	Is statement (2) ALONE enough?	Are statements (1) & (2) TOGETHER enough?	Appropriate choice or choices
1	YES	YES	--	D
2	YES	NO	--	A
3	NO	YES	--	B
4	NO	NO	YES	C
5	NO	NO	NO	E
6	YES	?	--	A or D
7	NO	NO	?	C or E
8	NO	?	--	B, C or E
9	?	YES	--	B or D
10	?	NO	--	A, C or E

Table 6

Independent Variables, Dependent Variables, and Covariates Used in the Multivariate Analysis of Covariance

Independent Variables	Dependent Variables (Posttest scores)	Covariates (Pretest scores)
1. Instructional group	1. QC geometry	1. QC geometry
a. QC	2. QC nongeometry	2. QC nongeometry
b. DS	3. DS geometry	3. DS geometry
c. RM	4. DS nongeometry	4. DS nongeometry
d. Control	5. RM geometry	5. RM geometry
2. Sex	6. RM nongeometry	6. RM nongeometry
a. Male		7. SAT-V
b. Female		

Table 7

Results of Multivariate Analysis
of Covariance for Main Effects
Due to Treatment

	Treatment effect		
	1st Canonical Variate (CV_1)	2nd Canonical Variate (CV_2)	3rd Canonical Variate (CV_3)
df hypothesis	18.00	10.00	4.00
df error	1363.79	965.00	483.00
F	12.372	6.269	5.799
$p <$	.001	.001	.001
R_c^2	.26	.08	.05

Correlations between dependent variables and canonical variates

Subtest	r_{cv_1}	r_{cv_2}	r_{cv_3}
QC geometry	.52	.22	.29
QC nongeometry	.82	.16	.01
DS geometry	.14	.49	.33
DS nongeometry	-.10	.81	.33
RM geometry	.21	-.20	.69
RM nongeometry	.03	-.21	.67

Mean discriminant scores

Group			
QC instructed	.82	-.02	-.18
DS instructed	-.39	.58	.05
RM instructed	.06	-.32	.37
Control	-.48	-.24	-.24

Table 8

Summary of Multivariate and Univariate Probability Levels Associated with Six Orthogonal Contrasts

Contrast	Multivariate Analysis	Univariate analyses					
		QC items		DS items		RM items	
		Geom	NGeom	Geom	NGeom	Geom	NGeom
QC - C	.001	.001	.001	.011	NS [a]	.012	.012
DS - C	.001	.001	.001	.001	.001	NS	NS
RM - C	.001	.001	.001	.025	NS	.001	.001
DS - RM	.001	NS	NS	NS	.001	.002	.006
QC - RM	.001	.001	.001	NS	NS	NS	.016
QC - DS	.001	.001	.001	NS	.001	NS	NS

[a] Not significant ($p > .05$).

Table 9

Results of Multivariate Analysis
of Covariance for Main Effects
Due to Sex

	Sex effect
df hypothesis	6.00
df error	482.00
F	3.140
$p <$	.005
R_c^2	.038

Correlations between dependent variables
and canonical variates

Subtest	r_{cv}
RM geometry	.41
RM nongeometry	.31
DS geometry	.64
DS nongeometry	.71
QC geometry	.37
QC nongeometry	.47

Mean discriminant scores

Sex	Discriminant Scores
Male	.22
Female	-.22

Table 10

Means and Standard Deviations for the QC, DS, and RM Scores, Standardized to a Pretest Scale with a Mean of 50 and Standard Deviation of 10

Subjects			Test											
			QC (55 items)				DS (43 items)				RM (42 items)			
Instructional Group	Sex	N	Pretest Mean	Pretest SD	Posttest Mean	Posttest SD	Pretest Mean	Pretest SD	Posttest Mean	Posttest SD	Pretest Mean	Pretest SD	Posttest Mean	Posttest SD
QC	M	74	52.0	10.0	61.4	10.8	51.7	10.9	57.9	12.6	52.0	10.4	54.8	11.9
	F	96	50.6	9.6	59.8	9.5	50.2	9.6	53.7	10.5	49.9	9.3	52.3	10.3
DS	M	28	50.5	12.6	55.4	11.8	50.5	12.7	62.1	16.0	51.1	11.4	53.4	11.7
	F	60	46.8	8.0	51.2	9.2	49.0	8.6	55.4	10.5	47.7	9.1	49.0	8.7
RM	M	34	51.2	11.7	58.0	10.8	51.9	11.3	57.2	14.0	52.1	9.4	57.2	10.3
	F	45	51.1	9.9	55.5	10.1	49.1	9.3	53.8	10.6	51.9	10.8	56.2	10.7
Control	M	57	52.7	11.0	55.7	11.6	51.7	11.1	55.8	9.5	52.6	10.2	54.5	10.7
	F	108	47.5	8.8	49.2	8.8	47.9	8.7	50.4	8.5	47.0	9.5	48.8	10.6

Means and standard deviations for each of the QC, DS, and RM tests have been standardized to the pretest scores with a mean of 50 and a standard deviation of 10. Thus a posttest entry represents the mean and standard deviation of a group in pretest units.

Table 11

Pretest Means and SDs and Mean Score Changes
(from Pretest to Posttest and to Post-posttest) on the
SAT-V and SAT-M for 417 Ss Who Took Three SATs

		SAT-V				SAT-M			
		Pretest		Mean change		Pretest		Mean change	
Group	N	Mean	SD	Oct. to Dec.	Oct. to Apr.	Mean	SD	Oct. to Dec.	Oct. to Apr.
QC	145	405	85	19	39	415	81	29	57
DS	72	390	79	12	34	401	75	37	52
RM	71	398	79	14	34	426	78	43	68
Control	129	387	87	8	28	392	74	18	47

Table 12

Selected QC Items, with Pretest and Posttest Item Data

Directions: Blacken space

A if the quantity in Column A is the greater;
B if the quantity in Column B is the greater;
C if the two quantities are equal;
D if the relationship cannot be determined from the information given.

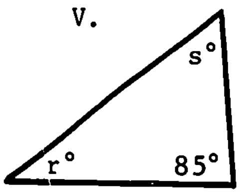
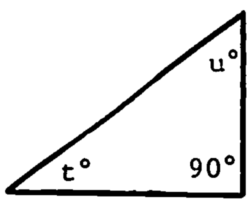
Item	Column A	Column B	Time of test	Delta	A	B	C	D	Omit
I.	3 X 268 X 8	2 X 268 X 9	Pre	12.0	54	7	28	1	0
			Post	8.1	76	4	5	0	0
II.	$\frac{x}{y}$	$\frac{x}{y} \cdot \frac{y}{x}$	Pre	14.3	19	30	5	33	3
			Post	9.9	6	10	2	66	1
III.	$\frac{x^2y^2z^2}{x^3y^3z^3}$	$\frac{xyz}{3}$	Pre	16.0	25	29	4	20	8
			Post	11.6	13	11	1	52	4
IV.	$-1 < x < 1$, $-1 < y < 0$; x	y	Pre	15.9	49	5	5	18	5
			Post	12.4	23	4	2	43	3
V.	 r + s	 t + u	Pre	12.8	47	29	5	8	1
			Post	9.3	70	9	1	4	1
VI.	$-10 < x < -1$; $\frac{1}{x^5}$	$\frac{1}{x^4}$	Pre	13.0	18	45	1	12	14
			Post	8.5	2	74	0	5	4
VII.	x + y	−x −y	Pre	15.5	43	4	15	23	3
			Post	11.4	17	1	8	55	1
VIII.	$x^2 = 4$, $y^2 = 9$; x	y	Pre	17.2	0	57	3	8	0
			Post	10.5	0	17	0	50	0

Table 13

Selected DS Items, with Pretest and Posttest Item Data

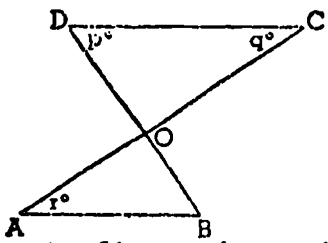
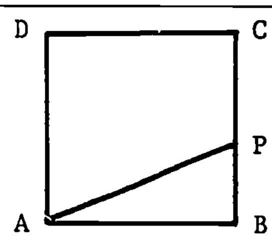
DS items	Time of test	Delta	A	B	C	D	E	Omit
			Response frequency					
I. What is the positive integer of which Bill is thinking? (1) Its square is between 20 and 50.	Pre	13.6	7	1	20	2	15	0
(2) It is an even number.	Post	10.3	1	1	30	0	8	0
II.  In the figure above, if AC and BD are straight lines intersecting at O, what is the numerical value of r? (1) p = 60	Pre	17.0	0	11	18	3	7	4
(2) q = 40	Post	14.3	3	7	11	2	15	1
III. What is the value of the odd integer x? (1) 18 < 2x < 30	Pre	16.7	3	2	8	9	18	5
(2) 22 < 2x < 50	Post	13.0	1	1	20	4	14	0
IV.  In the figure above, ABCD is a square, and point P is on side BC. What is the length of PC? (1) PC = 8 - PB	Pre	18.4	8	4	4	9	15	5
(2) AP = 2PC	Post	15.7	4	5	10	4	1[illegible]	3

Table 13, Continued

Selected DS Items, with Pretest and Posttest Item Data

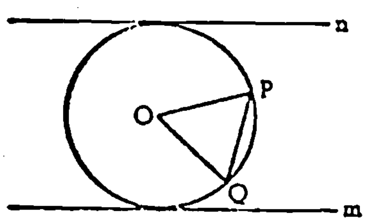
DS items	Pretest and posttest item data							
	Time of test	Delta	Response frequency					
			A	B	C	D	E	Omit
V. How much did John earn from the sale of 60 magazine subscriptions? (1) Each subscription sold for $2.00.	Pre	12.4	8	3	25	5	2	0
(2) He received a 20 percent commission on each sale.	Post	7.1	2	0	37	1	0	0
VI. How many dollars did Susie spend for clothing material? (1) The material cost 50 cents per square yard. (2) If she had bought twice as much material, she	Pre	12.4	8	5	4	1	25	0
could have made three more dresses.	Post	6.4	0	0	2	0	38	0
VII. If x is a member of the set of numbers (9, 12, 14, 35, 36), what number is x?	Pre	16.2	1	4	10	4	22	2
(1) x is a multiple of 2. (2) x is a multiple of 7.	Post	11.5	1	3	26	2	7	1
VIII.  In the figure above, the circle with center O touches two parallel lines, m and n and P and Q are on the circle. What is the distance in inches between m and n? (1) ΔOPQ is equilateral.	Pre	16.8	3	9	8	2	14	8
(2) The length of PQ is 2 inches.	Post	12.7	1	1	21	0	15	2

Figure Captions

Figure 1. A decision tree showing the relationships among choices for DS items under conditions of full, partial, or no information.

Figure 2a. Scatterplot of pretest and posttest QC item deltas, based on the responses of Ss instructed for QC (N = 85).

Figure 2b. Scatterplot of pretest and posttest DS item deltas, based on the responses of Ss instructed for DS (N = 40).

Figure 2c. Scatterplot of pretest and posttest RM item deltas, based on the responses of Ss instructed for RM (N = 40).

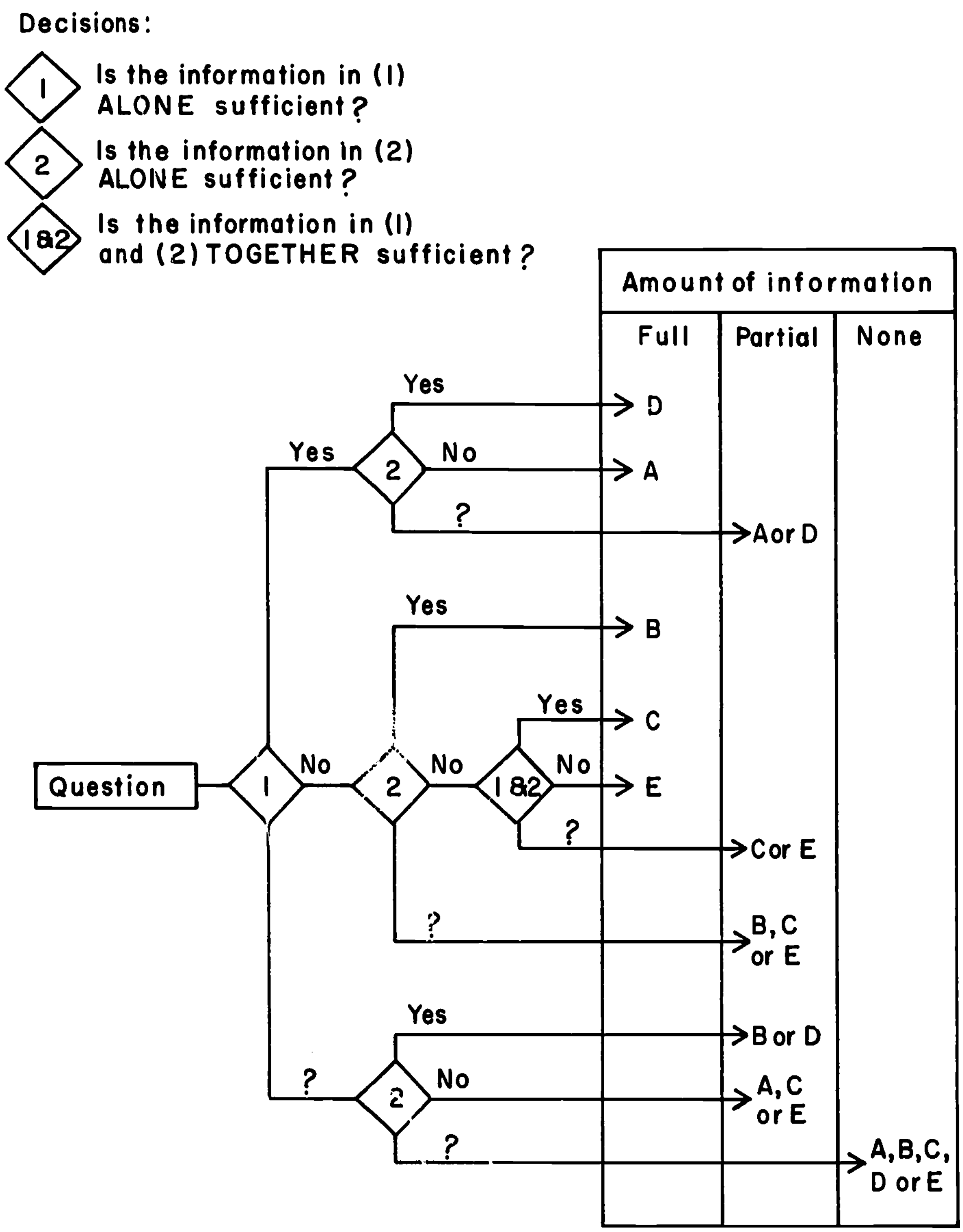
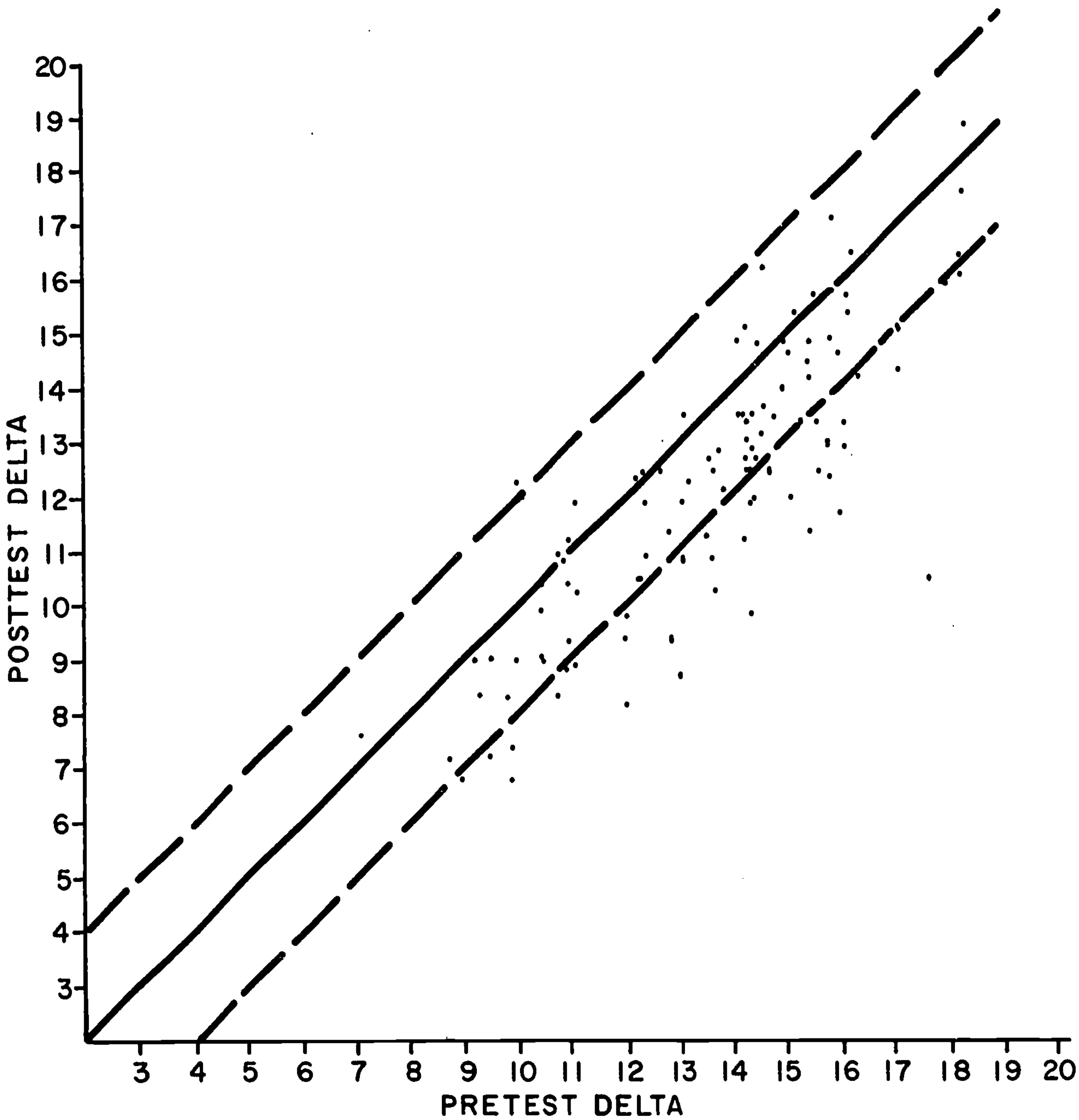

Decisions:
1
Is the information in (1) ALONE sufficient?
2
Is the information in (2) ALONE sufficient?
1&2
Is the information in (1) and (2) TOGETHER sufficient?
Amount of information
Full
Partial
None
Question
Yes
No
?
D
A
A or D
B
C
E
C or E
B, C or E
B or D
A, C or E
A, B, C, D or E

Figure 1



POSTTEST DELTA
20
19
18
17
16
15
14
13
12
11
10
9
8
7
6
5
4
3
3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
PRETEST DELTA

Figure 2a

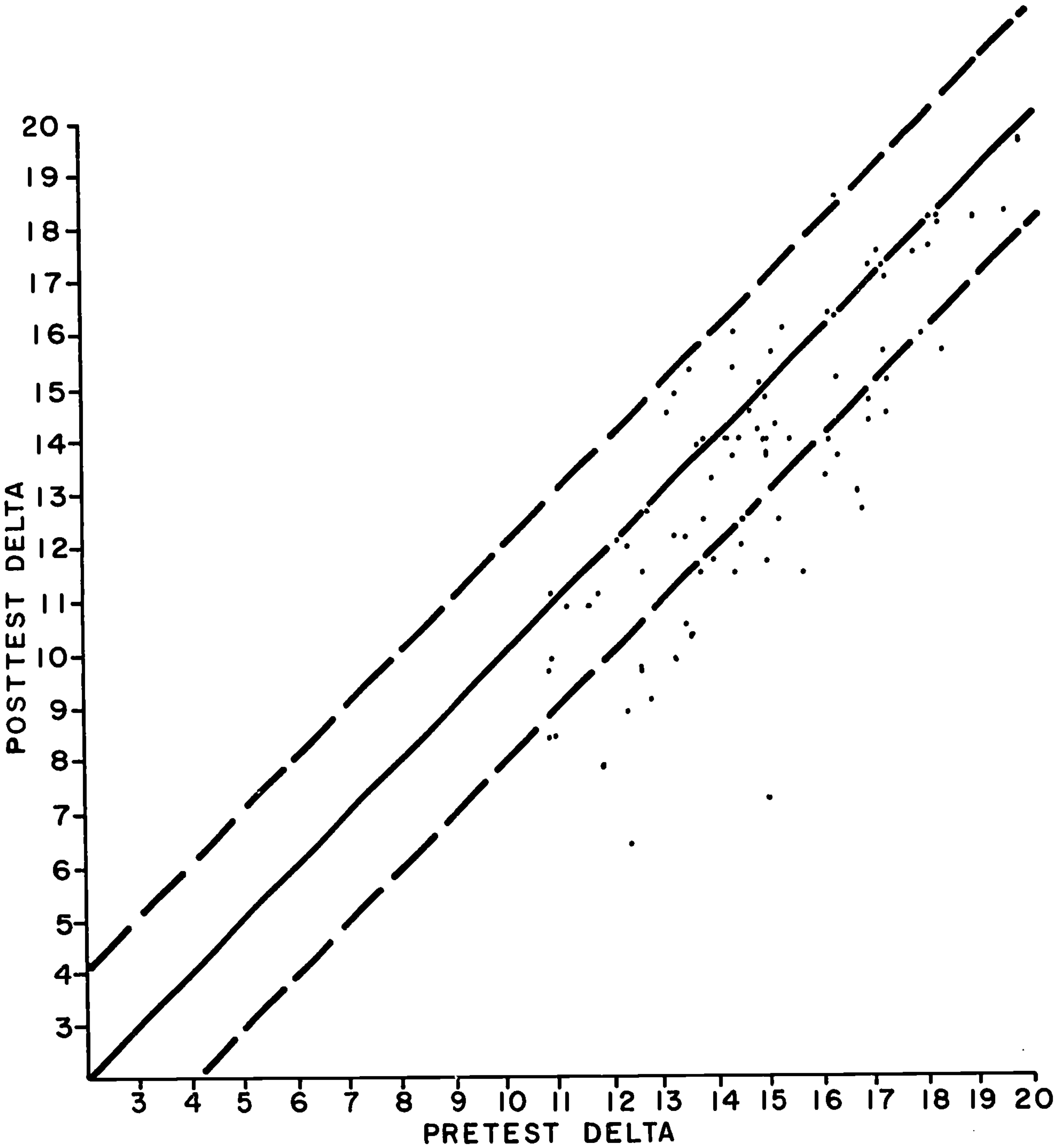

POSTTEST DELTA
20
19
18
17
16
15
14
13
12
11
10
9
8
7
6
5
4
3
3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
PRETEST DELTA

Figure 2b

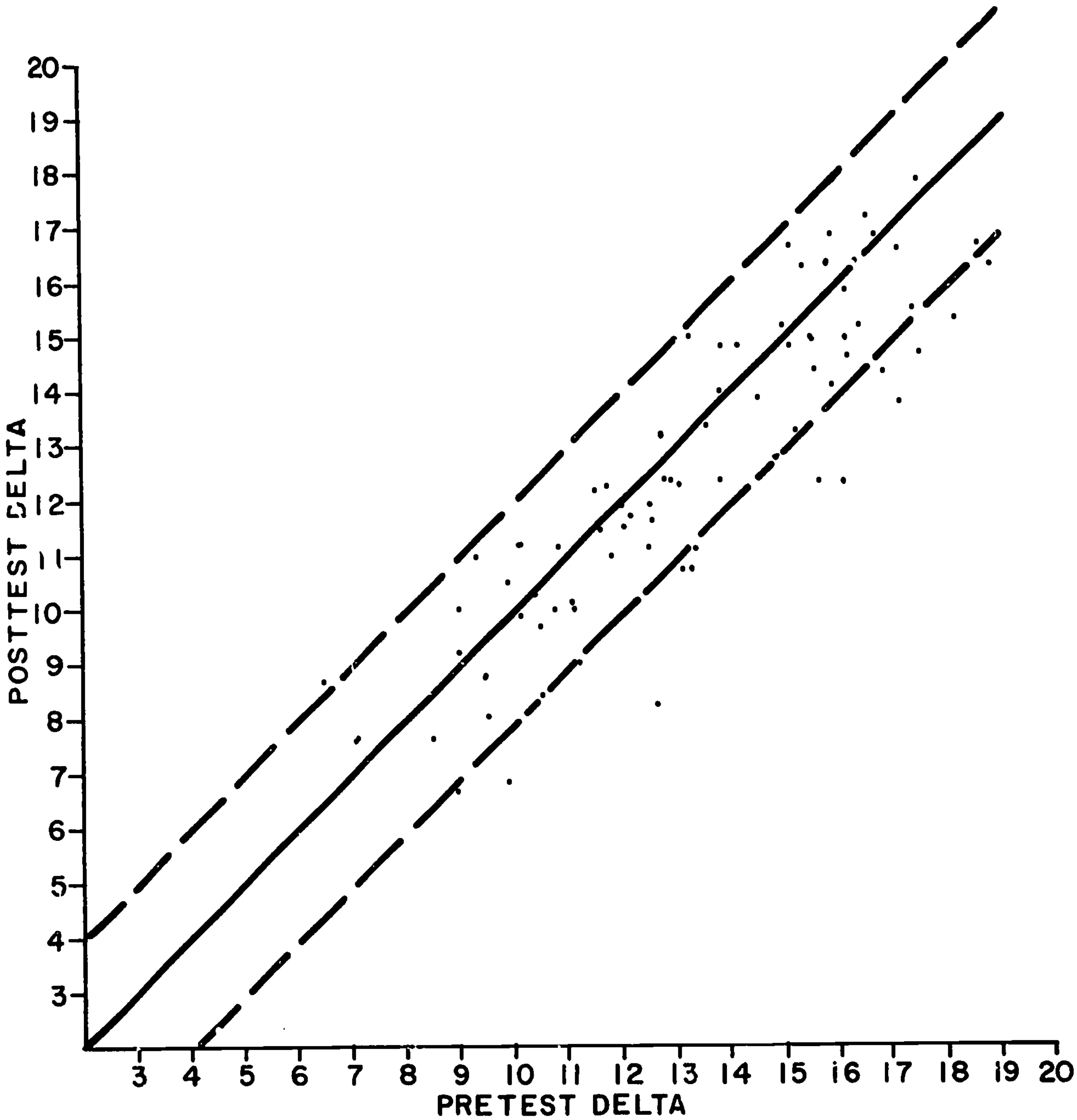

POSTTEST DELTA
20
19
18
17
16
15
14
13
12
11
10
9
8
7
6
5
4
3
3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
PRETEST DELTA

Figure 2c

Appendix A

Letters Sent to Prospective Subjects and Their Parents

EDUCATIONAL TESTING SERVICE PRINCETON, N.J. 08540

Area Code 609
921-9000
CABLE-EDUCTESTSVC

Developmental Research Division

September 23, 1970

Dear Parent:

Your child has volunteered to receive (free of charge) special instruction to prepare him for the mathematics section of the Scholastic Aptitude Test--the SAT. This instruction will be offered as part of a research project conducted by the Educational Testing Service and the College Entrance Examination Board. The project is designed to test whether special instruction can noticeably improve students' scores on the SAT. It is part of a continuing effort by Educational Testing Service and the College Entrance Examination Board to assure fairness to all students who take the examination.

Because of the nature of the research project, not all of the students who volunteer can be included in the program. Within the next three weeks, both you and your child will be notified of whether he is to be included in the project.

If your child is selected, he must be able to accept the following responsibilities:

(1) to come to the high school every Saturday morning, from October 17 through December 19 (except for November 28th, when there will be no instruction). Instruction begins at 9 a.m. and ends at noon.

(2) to do 2 to 3 hours of homework each week that will be assigned at the Saturday morning classes.

(3) to be responsible for his own transportation to and from the high school each Saturday morning.

If your child is chosen to participate in this study, he will receive instruction which has been designed and will be closely monitored by ETS staff. He will also be able to take a regular Scholastic Aptitude Test (SAT) in January or in April 1971, free of charge, and have his scores reported to three colleges of his choice.

We at ETS and at the College Entrance Examination Board are appreciative that your child volunteered for this project. We are looking forward to your interest and support.

Sincerely yours,

Franklin R. Evans and Lewis W. Pike
Project Directors

Appendix A, continued

EDUCATIONAL TESTING SERVICE PRINCETON, N.J. 08540

Area Code 609
921-9000
CABLE-EDUCTESTSVC

Developmental Research Division

Registration Number

Dear

This is to inform you that you have been selected to participate in the research project that is designed to test whether special instruction can affect students' scores on the math section of the Scholastic Aptitude Test (SAT).

Please remember that in order for the study to be meaningful it is important that you:

1. come to the high school every Saturday morning from October 17 through December 19 (except for November 28, when there will be no instruction).

2. do 2 to 3 hours of homework each week that will be assigned at the Saturday morning classes.

3. provide your own transportation to and from the high school each Saturday morning.

Remember also, that you will be taking a special administration of the SAT on Saturday, October 17. Be sure to arrive on time. Bring this letter which has your Registration number in the upper right-hand corner. Also be sure to bring 3 sharpened number 2 pencils. No other papers or books will be allowed in the classrooms used for testing.

The PSAT will be administered during the week after you have taken the SAT on October 17. Since the PSAT is a two-hour version of the SAT, measures the same mathematical and verbal abilities, and is in a very real sense practice for the SAT, you should not take the PSAT. The school will refund your money if you have already registered for the PSAT. You will be given your scores on the SAT that you take on October 17 after you take the second SAT on December 19. You may share these scores with your guidance counselor if you wish. You will also be able to take a regularly scheduled SAT in January, March or April 1971 free of charge and have your scores reported to three colleges of your choice.

This letter will serve as your ticket of admission to the special testing session on October 17. Please take care of it and bring it to school with you on that day.

We thank you for volunteering for this project and hope that it will be an interesting and profitable experience for you.

Sincerely yours,

Franklin Evans and Lewis Pike
Project Directors

Appendix A, continued

EDUCATIONAL TESTING SERVICE PRINCETON, N.J. 08540

Area Code 609
921 - 9000
CABLE-EDUCTESTSVC
•
Developmental Research Division

Registration Number

Dear

This is to inform you that you have been selected to participate in the research project that is designed to test whether special instruction can noticeably improve students' scores on the math section of the SAT (Scholastic Aptitude Test).

You have been placed in the group of students who will receive instruction after the Christmas holidays. The time and place will be mutually agreed upon later by the students in your group and your teacher. However, you must come to both of the special administrations of the SAT which are a part of the study. The first is on October 17 and the second is on December 19. Your attendance at these testing sessions is essential. Be sure to arrive on time. Bring this letter which has your registration number in the upper right-hand corner. Also be sure to bring 3 sharpened number 2 pencils. No other papers or books will be allowed in the classrooms used for testing.

The PSAT will be administered during the week after you have taken the SAT on October 17. Since the PSAT is a two-hour version of the SAT, measures the same mathematical and verbal abilities, and is in a very real sense practice for the SAT, you should not take the PSAT. The school will refund your money if you have already registered for the PSAT. You will be given your scores on the SAT that you take on October 17 after the second SAT you take on December 19. You may share these scores with your guidance counselor if you wish. You will also be able to take a regularly scheduled SAT in January, March or April 1971 free of charge and have your scores reported to three colleges of your choice.

This letter will serve as your ticket of admission to the special testing session on October 17. Please take care of it and bring it to school with you on that day.

We thank you for volunteering for this important project and hope that it will be an interesting and profitable experience for you.

Sincerely,

Lewis W. Pike and Franklin R. Evans
Project Directors

Appendix B: Examples of Mini-Lessons

Mini-Lesson 1

MEMORIZE THE BASIC FORMULAS

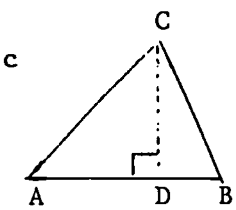
At the beginning of the test there is always some information that you can use on the test. YOU ARE SURE TO NEED THIS INFORMATION, so you should MEMORIZE IT.

The following information is for your reference in solving some of the problems:

Circle of radius r:

Area = πr^2
Circumference = $2\pi r$
The number of degrees of arc in a circle is 360.

The measure in degrees of a straight angle is 180.



Triangle:

The sum of the measures in degrees of the angles of a triangle is 180.

If $\angle CDA$ is a right angle, then

(1) area of $\Delta ABC = \frac{AB \times CD}{2}$

(2) $AC^2 = AD^2 + DC^2$

Definitions of symbols:

$<$ is less than
$>$ is greater than
$\perp$ is perpendicular to
$\leqq$ is less than or equal to
$\geqq$ is greater than or equal to
$\parallel$ is parallel to

Mini-Lesson 6

MATH VOCABULARY SHOULD BE AT YOUR FINGERTIPS

When you see:	You should think:
CONSECUTIVE INTEGERS	n, n+1, n+2, n+3, n+4, ...; where n is an integer
CONSECUTIVE EVEN INTEGERS	2n, 2n+2, 2n+4, 2n+6, ...; where n is an integer
CONSECUTIVE ODD INTEGERS	2n+1, 2n+3, 2n+5, 2n+7, ...; where n is an integer
AVERAGE	ADD the terms; COUNT the terms; DIVIDE

$$\text{AVERAGE} = \frac{\text{SUM of terms}}{\text{NUMBER of terms}}$$

For example:

$$\text{AVERAGE of 9, 11, and 16} = \frac{9 + 11 + 16}{3} = 12$$

Appendix C

Examples of Workbook Exercises Involving Mathematics Content

An exercise for using information from geometric figures is shown below. Answers are on the following page. Also on that page is an exercise that was used for teaching inequalities.

- -

Directions:

It is important to know what kinds of questions can be answered when information is given using geometric figures. It is also important, of course, to be able to answer those questions for which the necessary information is provided.

There are 5 problems in this exercise. Each problem includes a labeled geometric figure and some information about that figure. For each problem, 5 questions are asked (a,b,c,d, and e). For each of these questions, decide first whether it could be answered, using the information provided, and indicate by writing "yes" or "no" in the first blank. (Enough information) Then, if you answered "yes," try to answer the question and write the answer in the second blank. (Answer)

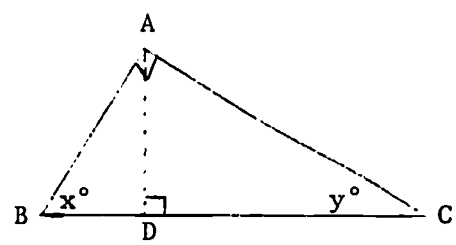
Answers and explanations for each problem are given on the page following the problem.

NOTE 1: Often a particular question can be answered, with the help of answers to earlier questions in the same problem.

NOTE 2: It would be a good idea to review the mini-lesson cards referring to geometric figures, before doing this exercise.

- -

Problem 1.



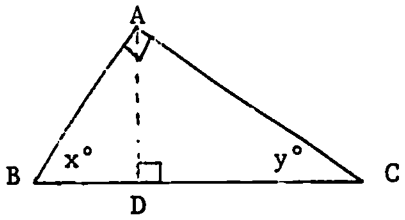
$x = y$ and $AB = 2$

	Enough Information?	Answer
a. What is the value of x?	______	______
b. What is the length of AC?	______	______
c. What is the area of ΔABC?	______	______
d. What is the length of BC?	______	______
e. What is the length of AD?	______	______

Appendix C, continued

Answers to

Problem 1.



$x = y$ and $AB = 2$

Question	Enough Information?	Answer	Explanation
a. value of x	yes	45	Angle A = 90°, so $x + y = 90$; $x = y$, so $x = 45$
b. length of AC	yes	2	Since $x = y$, $AC = AB = 2$
c. area of ΔABC	yes	2	Area = $\frac{1}{2}$ (AB)(AC)
d. length of BC	yes	$2\sqrt{2}$	$BC^2 = AB^2 + AC^2 = 4 + 4$. $BC = \sqrt{8} = 2\sqrt{2}$
e. length of AD	yes	$\sqrt{2}$	AD is an altitude of ΔABC. Area = $\frac{1}{2}$(base)(altitude), so altitude = $\frac{2(\text{area})}{BC} = \frac{4}{2\sqrt{2}} = \sqrt{2}$ There are other ways to determine AD.

An inequalities exercise

1	$\frac{2}{3}$	$\frac{3}{5}$	$\frac{1}{2}$	$\frac{2}{5}$	$\frac{1}{4}$	$\frac{3}{8}$	0	-1	$-\frac{2}{3}$	$-\frac{3}{5}$	$-\frac{1}{2}$	$-\frac{2}{5}$	$-\frac{1}{4}$	$-\frac{3}{8}$

1. Which of the above could be a if $2 < 5a < 3$?
2. Which could be b if $\frac{1}{2}(b + \frac{1}{3}) < \frac{1}{2}$?
3. Which could be c if $\frac{1}{c^2} > 4$?
4. Which could be d if $0 < d + \frac{1}{d} < 3$?
5. Which could be e if $\frac{1}{e} > 2$?

Appendix D: Examples of the Explanations Provided for DS Items

---

Example 1. In the past 5 years, a lawyer won 89 of his court cases and lost the remainder. What percentage of his cases did he win?

(1) He was involved in 103 court cases during the past five years.

(2) He lost 14 cases during the past 5 years.

- -

(1) ALONE? YES. $\%\text{ won} = \dfrac{\text{number won}}{\text{number involved in}}$.

Remember Rule I. Do not actually compute the percentage! (Correct choice must be A or D.)

(2) ALONE? YES. Why? (Correct choice must be B or D.)

D is correct.

Example 2. Is it true that $x > 6$?

(1) $x > 4$

(2) $-x < -6$

- -

(1) ALONE? NO. We are only able to determine that $x > 4$. x could be 5, 6, 7, etc. For $x = 5$ and $x = 6$, it is obvious that $(x > 6)$ is not true and that for $x = 7$, etc., $x > 6$ is true. (Correct choice must be B, C or E.)

(2) ALONE? YES. The expression given in (2) is the same as $x > 6$. Multiplying both sides of an inequality by the same negative number changes the direction of that inequality. $-x < -6$ is the same as $x > 6$. (Correct choice mus be B or D.)

If it was easy for you to answer (1), but you were not sure of (2), did you use Rule III?

B is correct.

Appendix E, continued

Directions: For each problem, mark an X through one answer choice. Be sure to use the partial information that is provided.

Problem	(1) ALONE?	(2) ALONE?	(1) and (2) TOGETHER?	ANSWER CHOICE
1.	yes	?		1. A B C D E
2.	no	no	?	2. A B C D E
3.	?	yes		3. A B C D E
4.	?	no		4. A B C D E
5.	?	no		5. A B C D E
6.	no	no	?	6. A B C D E
7.	yes	?		7. A B C D E
8.	no	no	?	8. A B C D E
9.	yes	?		9. A B C D E
10.	no	?		10. A B C D E
11.	yes	?		11. A B C D E
12.	?	yes		12. A B C D E
13.	no	?		13. A B C D E
14.	no	no	?	14. A B C D E
15.	yes	?		15. A B C D E
16.	yes	?		16. A B C D E
17.	?	no		17. A B C D E
18.	?	yes		18. A B C D E
19.	yes	?		19. A B C D E
20.	no	no	?	20. A B C D E

- -

COMPUTATION OF TEST SCORE:

Number Right ________________(a)

4(Number Right) ________________(b) = 4(a)

Number Wrong ________________(c)

$$\frac{b - c}{4} = \frac{\qquad}{4} = \boxed{\qquad}$$ TEST SCORE 2

Appendix E

A Classroom Demonstration of the Effects of Guessing Under Conditions of No Information and of Partial Information for DS Items

---

A very simple exercise can illustrate the value of using partial information. The first part will illustrate the results of blind guessing (answering when there is no information). The second part will show what happens when students answer on the basis of partial information.

First, distribute Answer Sheet 1. (A copy of answer sheets 1 and 2 are appended to this outline.) Tell the students that the correct answer choices were taken from a real test. Their task is to "try their luck," by picking one answer for each problem. Ask them to answer quickly.

As soon as the students have completed marking answers, distribute Answer Sheet 2. Read the directions to them, then have them make their answers.

Then, have students mark Answer Sheet 1 (with red pencil), according to the following key:

Key for Answer Sheet 1 and Answer Sheet 2

1. D	6. C	11. A	16. D
2. C	7. A	12. B	17. C
3. B	8. E	13. C	18. D
4. E	9. D	14. C	19. A
5. C	10. E	15. A	20. E

Next, have the students compute TEST SCORE 1. Get a rough tally of these scores, and find out highest and lowest scores. Most scores should be near zero, and the average should be very near zero.

Point out that on the average blind guessing has no effect on their scores. For this group of items, the score when every item is attempted is likely to be the same as the score if they did not answer any of the items. (That is, a "zero" score, either way.)

Next, have students mark Answer Sheet 2, using the same key used for Answer Sheet 1, and have them compute TEST SCORE 2. Again, get a tally of the scores. This time, scores should average around 6 or 7. Point out that by using partial information properly, most students have gained about 6 or 7 score points, by guessing on 20 items they might otherwise have skipped over. (6 or 7 more items correct on an SAT would result in a sizeable gain in SAT score.)

Answer Sheet 2 is shown on the next page. Answer Sheet 1 was identical, except that no sufficiency information was given, and the directions read: "For each problem, mark an X through one answer choice."

Appendix F

Examples of Test Item Explanations in QC Workbooks

Example 1:

Column A	Column B
$(37)\left(\frac{1}{43}\right)(58)$	$(59)\left(\frac{1}{43}\right)(37)$

The answer is B

(Rule I says it cannot be D)

Remember rule III and eliminate common terms

$\frac{1}{43}$ and 37 from both sides

This leaves

$$58 \quad < \quad 59$$

Example 2:

Column A		Column B
x^2	$x < y$	y^2

The answer is D. Rule II

Don't forget negatives!

if $x = -2$ and $y = 1$ then

$$x^2 = 4 \quad > \quad y^2 = 1$$

but if $x = 1$ and $y = 2$ then

$$x^2 = 1 \quad < \quad y^2 = 4$$

therefore

$$x^2 \quad ? \quad y^2$$