DOCUMENT RESUME

ED 383 774 TM 023 947

AUTHOR	Braswell, James S.; Jackson, Carol A.
TITLE	An Introduction of a New Free-Response Item Type in Mathematics.
PUB DATE	19 Apr 95
NOTE	31p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (San Francisco, CA, April 19-21, 1995).
PUB TYPE	Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	*Computer Assisted Testing; Educational Assessment; Guessing (Tests); *Mathematics Tests; Multiple Choice Tests; *Problem Solving; Responses; Scoring; Student Attitudes; *Student Evaluation; Teacher Attitudes; Test Construction; Test Format; *Test Items
IDENTIFIERS	*Free Response Test Items; Preliminary Scholastic Aptitude Test; Scholastic Aptitude Test

ABSTRACT

A new free-response item type for mathematics tests is described. The item type, referred to as the Student-Produced Response (SPR), was first introduced into the Preliminary Scholastic Aptitude Test/National Merit Scholarship Qualifying Test in 1993 and into the Scholastic Aptitude Test in 1994. Students solve a problem and record the answer by blackening the ovals on a grid that permits numbers from 0 to 999. For the test taker, the use of the SPR format provides a more natural problem-solving situation in which the student must analyze and solve the problem without being influenced by multiple choice alternatives. Responses from students and teachers have generally been positive, agreeing that the format reflects the mathematical ability of the student better than multiple choice items. There are some drawbacks to the approach, including the length of time required to grid the answer. The ability to enter answers directly into a computer will eliminate this difficulty. The reduction or elimination of guessing has resulted in items with better discrimination indices and improved test reliability. Two figures and two sample items illustrate the discussion. Appendixes A and B contain directions and a sample score report; Appendixes C and D present sample items. (SLD)

AN INTRODUCTION OF A NEW FREE-RESPONSE ITEM TYPE IN MATHEMATICS

by

James S. Braswell

&

Carol A. Jackson

Educational Testing Service

Paper presented at the annual meeting of the National Council on Measurement in Education
San Francisco, April 19, 1995

AN INTRODUCTION OF A NEW FREE-RESPONSE ITEM TYPE IN MATHEMATICS

Introduction

Standardized examinations in most high-volume testing programs continue to use multiple-choice test questions, primarily for reasons of efficiency and economy. Over the years the multiple-choice questions have proven to have high reliability and reasonable predictive validity. This format has been used to pose questions covering a range of thinking skills -- from routine type problems to those involving higher order thinking skills. However, with some justification, multiple-choice questions have been criticized for rewarding recognition rather than the production of answers and for possibly being susceptible to short-term score gain through the application of test-wise strategies.

The purpose of this paper is to describe a new free-response item type that was first introduced into the PSAT/NMSQT program in the fall of 1993 and into the SAT I program in the spring of 1994. The paper will highlight advantages and disadvantages of the new item type, discuss how responses are edited and scored, and present sample questions with a discussion of the rich variety of responses given by students.

Background

During the period 1988-1992 Educational Testing Service and the College Board undertook studies to determine the feasibility of including non-multiple-choice mathematics questions on the SAT I and the PSAT/NMSQT. The goal was to have some mathematics questions that could be machine-scored yet not presented as traditional multiple-choice questions. Early investigations used a 3-digit integer grid in field trial studies for students to record their answers. Using this grid (see Figure 1), it was possible for students to solve a problem with an integer answer from 0-999 and enter the answer on a grid that could be machine-scored.

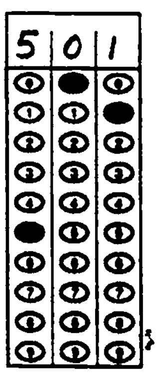
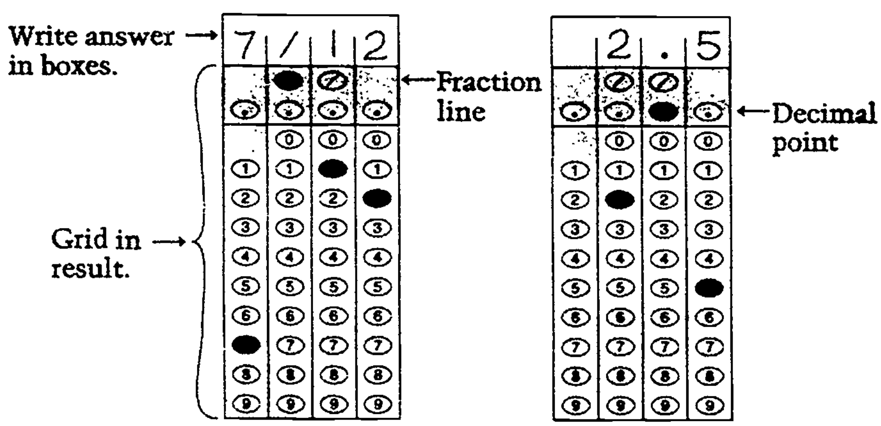


Figure 1

In entering answers on the grid, students were asked to write their solution at the top of the grid and then darken the corresponding ovals. Prior research related to gridding other types of information such as a street address revealed that greater accuracy is achieved when a result is written before it is gridded.

The 3-digit integer grid format worked reasonably well, but seemed too limited since it did not allow for the possibility of fraction and decimal answers. Also, the statistical data (item difficulty and biserial correlation) was not appreciably different from data derived from the administration of similar multiple-choice items. A subsequent investigation considered the possibility of using a more elaborate grid that would allow for fractions, decimals, integers from 0-9999, and negative numbers. To avoid overly complicating matters, a decision was made not to include negative numbers on the grid and also not to allow for the possibility of variables such as x or y. Eventually, the grid evolved to its current operational form shown in Figure 2. Complete directions for answering this item type, which is now referred to as Student-Produced Response (SPR), are given in Appendix A.

Figure 2



Implementation

In October 1993 the SPR item type was introduced in the Tuesday and Saturday forms of the PSAT/NMSQT. Approximately 1.8 million students took these two forms and each form contained 10 SPR questions. This new PSAT/NMSQT contained a total of 50 mathematics questions. In March 1994, the SPR item type was introduced in the SAT I, the first administration of the revised SAT. This test form also contained 10 SPR questions. This new SAT I contained a total of 60 mathematics questions.

The decision to use 10 items in the SPR format was a practical one based on available testing time, the fit with other item types, and answer sheet design considerations. Staff also thought that including fewer than 10 SPR items would not be worth the effort, and including significantly more than 10 would have involved too much risk when other changes were being introduced in the new PSAT/NMSQT and SAT I at the same time.

The scoring and reporting of results on the new tests in general, and the SPR items in particular, has been quite successful. For the PSAT/NMSQT the student's score report contains the correct answer(s) for each SPR question, the student's answer, and an indication as to whether the student's answer was correct. A copy of a sample score report is given in Appendix B. Copies of the Tuesday and Saturday forms of the PSAT/NMSQT are released shortly after their administration so students can check the results indicated on their score report against the actual test questions. Although not all forms of the SAT I are released, for the 2 or 3 tests per year that are released, students can obtain information similar to that made available for the PSAT/NMSQT.

Why Use SPR Items?

For the test-taker, the use of the SPR format provides a more natural problem-solving situation in which the student must analyze and solve a problem without being influenced by multiple-choice alternatives. Because guessing the answer is almost impossible, the SPR format discourages coaching strategies that have little long-term benefit. For example, students cannot work backwards from the choices to see which one "fits" the given situation. The absence of guessing as a significant factor enhances the reliability of the statistical data and gives this format greater face validity.

From the test-taker's perspective, the SPR format allows students to answer a question in a way that is consistent with their solution. For example, in a simple probability question some students might arrive at an answer of 3/12, others 1/4, and yet others .25. The grid in Figure 2 will accommodate all of these solutions and other equivalent solutions that will fit in the grid.

Yet another advantage of the SPR format is that it avoids the subtle hints that multiple-choice alternatives sometimes provide in a problem. For example, if a question involves finding the units digit of 3^{20} and the choices are

(A) 1 (B) 3 (C) 5 (D) 7 (E) 9

a student who does not understand the meaning of "the units digit" gets a clue from the choices. When this question was presented in the SPR format, over 20 percent of the responses were not digits. The fact that all of the digits in the answer choices are odd digits provides yet another clue to solving the problem. In the absence of choices, students solve problems and obtain answers that a test developer would sometimes never have considered in posing a multiple-choice question. (See Sample Items section for examples.)

The use of the SPR item type is also more consistent than the multiple-choice type with the recommendations of groups such as the National Council of Teachers of Mathematics (NCTM). The NCTM has published a set of standards for both curriculum and evaluation. The Evaluation Standards strongly encourage alternatives to traditional multiple-choice testing. Although this item format does not allow for longer problems for which students' written work is evaluated, it is a first step in introducing constructed response items in a large scale standardized testing program. In line with the Standards, the SPR format does provide for the opportunity of more than one correct answer. Currently the maximum number of discrete correct answers the system will accept is 6. The correct answer can also be any value in a specified interval (e.g., $\frac{1}{3} < x < \frac{2}{3}$). Examples 4 and 5 in Appendix D illustrate these two possibilities.

In general, reactions from various groups (teachers and students) both before and after the introduction of this format have been positive. Students have indicated that the format is more consistent with the way they solve problems in school -- i.e., without the presence of multiple-choice options. In surveys conducted during the field testing of this item format, students also indicated that this type of question better represented their mathematical ability than did standard multiple-choice questions.

Negative Aspects of the SPR Format

The SPR format does have some negatives. It takes somewhat more time to grid a SPR answer than to grid a multiple-choice alternative. Test takers have no chance to recover from minor errors and receive no immediate feedback about the correctness of their result as they would in the multiple-choice format. In the multiple-choice format, students who solve a problem and obtain one of the answer choices take some comfort in the fact that they may have the correct answer. However, the new PSAT/NMSQT and SAT I both allow the use of a calculator and this technology reduces the likelihood of purely mechanical (computational) errors as a factor.

The directions for answering SPR questions are also more complex and longer than for the standard multiple-choice. For example, students cannot enter a mixed number such as $2\frac{1}{2}$, but must convert it to 2.5 or 5/2. This gridding rule and others are included in the directions that are available for reference to the test taker before and during the test. Again, Appendix A shows the complete directions for the SPR questions and for the standard multiple choice questions.

From an operational point of view, the answer sheet for recording SPR answers is more complex than that of the standard multiple choice answer sheet and the scoring rules are non-trivial. For example, if the correct answer to a question is 2, every possible result that is equivalent to 2 must also be scored as correct (e.g., 2.0, 1/.5, 6/3). The fact that the scoring rules are complex, however, is an administrative issue and is not a problem for students. In Appendix C there is a sample SPR question with a complete listing of all possible correct answers. In this appendix, the column headed "Edited", applies the editing rules for SPR answers discussed in the next section.

Editing Rules

The answer sheets which collect SPR responses are designed with four columns per item (see Figure 2). Bubbles are provided for a decimal point (.), a fraction bar (/), and digits 0 through 9. The fraction bar only makes sense in columns 2 and 3 and therefore is only included in these columns. The digit 0 makes sense in all four columns, but is not included in column 1 because for some questions it could lead to less precise results -- for example, 0.66 rather than .666. In the directions, students are instructed to enter the most precise value the grid will accommodate. In fact, the .666 example is included in the directions shown in Appendix A.

Students are expected to grid their answers in one or more columns. It does not matter whether students left-justify or right-justify their answers. Columns without gridding are read as blanks. Columns with multiple grid marks of roughly equal intensity are identified by the scanner with a special symbol and ultimately scored as incorrect. If columns contain multiple-grid marks of different intensities, the scanner has sophisticated logic that will accept the darkest mark as the intended response. To facilitate scoring, students' gridded responses are edited. The purpose of editing is to increase the probability that students who have solved problems correctly will not lose credit because of a quirk in the way they entered their answer on the grid. The editing rules are quite formal and extensive, so only a few of them will be illustrated below. In the examples that follow, '_' will be used to indicate a blank column.

Example 1: The first step in the editing process is to remove all blanks.

a. _2_3 is edited to 23
b. ._5_ is edited to .5

Example 2: Reset // to / and reset .. to .

a. 2//3 is edited to 2/3
b. ...1 is edited to .1
c. .1.. is edited to .1. and is invalid (decimal point on both sides of digit)

Example 3: If a / is present, there must be a digit in both the numerator and the denominator in order to be a valid response.

a. 2/.5 is valid (and equivalent to 4)
b. 2/.. is edited to 2/. and is invalid since there is no digit in the denominator
c. If the denominator is 0, the response is invalid -- e.g., 21/0 is invalid

Some additional examples of item response editing.

Raw Responses	Edited Response
_2/9 2/_9 2//9	2/9 (valid response)
.23_ _.23 ..23	.23 (valid response)
_10_ _ _10 _1_0	10 (valid response)
/2/3 ./2/	/2/3 ./2/ (invalid response)
../2	./2 (invalid response)

Again, the primary purpose of editing is to "clean up" the gridded response and thereby increase the likelihood that fair and accurate scores are awarded.

It should be noted that a comprehensive gridding error study was done prior to the introduction of SPR items. Although some students made gridding errors of various types, these errors likely occurred for two primary reasons -- (1) students were not motivated as they would be in an operational setting where results count, and (2) limited materials were available to acquaint students with the gridding rules. Information derived from the gridding error study is reflected in the final version of the test directions and also in the "tips" given in the publications which students receive before they take the PSAT/NMSQT or the SAT I.

Item Response Data for SPR Items

A special computer program is used to analyze the responses to the SPR questions. This program will produce a "most popular responses list" and a "high ability list." For PSAT/NMSQT, the most popular responses list is based on the first 25,000 answer sheets for each test form. Fifty of the most popular responses to each item are provided. The high ability list includes all unique ways the high ability examinees responded to an item. The high ability population for a 50-item PSAT/NMSQT test consists of only those who scored at or above a formula score (score adjusted for guessing) of 35 on the entire mathematics portion of the test.

For SAT I, the most popular responses list is based on a spaced random sample of juniors and seniors who took a particular subform of the SAT I test. There are approximately 3,500 answer sheets in this sample for each test. Up to 99 of the most popular responses to each item are provided. Included in the high ability population for an SAT I test are only those who scored at or above a scaled score of 660 on the mathematics portion of the test (the scaled score range is 200-800).

The SPR responses are edited and collapsed using the computer Equivalent Response Program (ERP). All edited and collapsed SPR responses are written to an output cartridge. ERP collapses any like responses according to previously defined rules. This output cartridge is then used as input to the program that produces the list of popular responses, thus eliminating redundant responses from the list of popular responses. For example, _1/4, 1/4_, .25_, 2/8, etc. are all collapsed to 1/4 once the ERP is run (the '_' is used to indicate a blank column on the grid).

The most popular response list as well as the high ability list are produced at both the pretest and final form stages to assist test developers in identifying potential problems in items with regard to both wording and keys. It is especially helpful at the pretest stage when questions can be flagged and revised before they are put into a final form.

Item Statistics for SPR Items

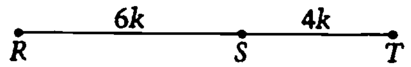
The observed-delta value used at ETS is a transformation of the proportion (p^+) of examinees answering an item correctly. Delta is equal to 13-4z, where z is the normal deviate corresponding to p^+. Delta is inversely related to p^+; i.e., the higher the delta the more difficult the item. For example, a high delta of 16.4 has a low p^+ of 20% correct, a middle delta of 13.0 has a middle p^+ of 50% correct, and a low delta of 7.9 has a high p^+ of 90% correct.

To put an item on scale, the observed deltas are then transformed to equated deltas by estimating the difficulty level of the item for the SAT standard-reference population. In this paper the delta values stated for the individual items will refer to the equated delta values.

Sample Items

Two SPR items, one from an SAT I test and the other from a PSAT/NMSQT test, are discussed in the following section. Appendix D includes an analysis of six other SPR items taken from these tests. *Careful analysis of the sample items together with related data reveals great variety in errant problem-solving approaches -- approaches that would not likely be used by test developers posing similar multiple-choice problems.*

Sample Item 1



If line segment RT above has length 5, what is the value of k ?

This question was on a disclosed SAT I test and had an equated delta of 10.4. Based on a sample size of approximately 3,500 students, approximately 2,500 gave the correct answer of 1/2 or .5. The next most popular response was to omit the question. In this sample, approximately 425 students omitted the question. The responses of 2, 10, 1, 2.5, 5, 1/5, and 3, listed in order of popularity, were given by at least 20 students. The interesting fact is that although the length of segment *RT* is 5, each of these seven responses for the value of *k*, except for 1/5, will obviously make the length of *RT* greater than 5.

If this question was presented in a multiple-choice format, probably most of the answer choices would have been less than 1. To ask this question and give as answer choices

(A) 0.5 (B) 1 (C) 2 (D) 2.5 (E) 10

(the correct answer with the four most popular responses given in the SPR format), would not have been considered good item construction. Quick inspection of these particular five answer choices would indicate that $k = 1$ makes the length of *RT* greater than 5, and since the answer choices are listed in increasing order (answer choices on both the PSAT/NMSQT and the SAT I are listed in either increasing or decreasing order), the values of *k* greater than 1 could also be eliminated from consideration. The only answer choice that makes sense among these five is (A) 0.5. It is considered good item construction practice to include reasonable answer choices that cannot be easily eliminated. An item whose answer can be easily guessed has little, if any, face validity.

The high ability list showed that in the sample only 4 high ability students answered 2, only 1 high ability student answered 1, and no high ability students gave other responses from the most popular responses list.

Sample Item 1

Test: SAT I
Equated Delta: 10.4
Sample Size: 3,500*

Response	Number of Students (Approximate)	Number of High Ability Students (Approximate)
Correct answer 1/2 or .5	2,500	400
Omits	425	---
2	125	4
10	70	---
1	60	1
2.5	60	---
5	40	---
1/5	20	---
3	20	---

*Note smaller sample size for SAT I tests compared to PSAT/NMSQT tests.

Sample Item 2

−980, −76, −54, 0, 1, 2, 3, 54, 76, 980

What is the average (arithmetic mean) of the 10 numbers in the list above?

This question was on a PSAT/NMSQT test and had an equated delta of 9.5. Based on a sample size of approximately 25,000 students, approximately 17,000 gave the correct answer of .6. The most popular wrong answer, answered by approximately 1,200 students, was 6 (the sum of the 10 numbers). Another wrong answer, also answered by approximately 1,200 students, was 1.5, obtained by either dividing 6 by 4 (since there are four numbers left after the other six numbers cancel each other out) or by giving the median as the answer. There were approximately 1,000 students who omitted the question and 500 students who answered 2 (6 divided by 3). Answers of 0, .5, 1, .4, 3, and .3, listed in order of popularity, were given by at least 100 students. The answer 3 shows a common misconception involved when computing an average -- to divide the sum by 2 no matter how many numbers are being averaged. If this question were asked in a multiple-choice format, answer choices of 1.5, 2, 3, and 1 would probably have been given with the correct answer of .6. Answer choices of 6, 0, .5, .4, or .3 may not have been considered as answer choices for a multiple-choice format of this question. Thus, in a multiple-choice version of this question, several common answers obtained by students would not have been included as choices, and students who obtained these common answers would have the opportunity to guess from among the five given choices. Guessing tends to make the test less reliable.

The high ability list showed that approximately 200 students in the high ability group answered 1.5, approximately 70 students in this group answered 2, and 45 in this group answered 6. The number of high ability students who gave other responses was much less than 45.

Sample Item 2

Test: PSAT/NMSQT
Equated Delta: 9.5
Sample Size: 25,000

Response	Number of Students (Approximate)	Number of High Ability Students (Approximate)
Correct answer .6	17,000	4,000
6	1,200	45
1.5	1,200	200
Omits	1,000	20
2	500	70
0	200	5
.5	200	11
1	150	8
.4	150	15
3	150	11
.3	140	13

Conclusion

Inclusion of the SPR item type on the PSAT/NMSQT and the SAT I has been a successful first step in including non-multiple choice questions in a high volume testing program. The SPR item type has given the test greater face validity among school and college faculty. Also, the fact that guessing and backdoor approaches have been reduced or eliminated has resulted in items with better discrimination indices and a test that has slightly higher reliability than its predecessor.

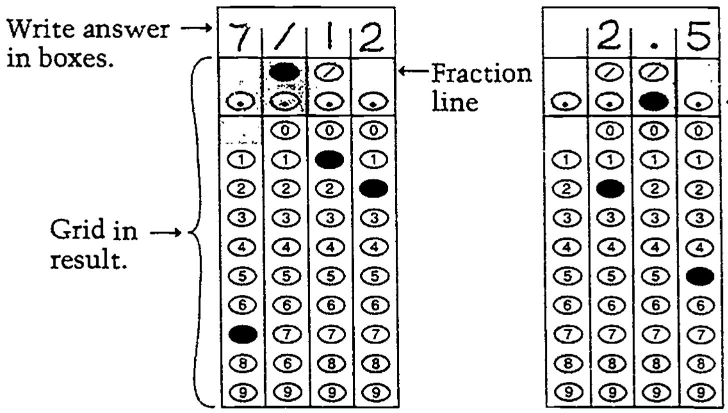
What has been learned from the successful implementation of this item type can be applied to the future development of computer-delivered versions of the SAT I test. Once the tests are in computer-delivered form, test takers will be able to enter their answers directly into the computer, thereby eliminating the need for the grid. For a computer-delivered test, it is also reasonable to consider the possibility of awarding partial credit for answers that reflect partial understanding. Experience gained in implementing the SPR item type has moved these ideas closer to the realm of possibility.

APPENDIX A

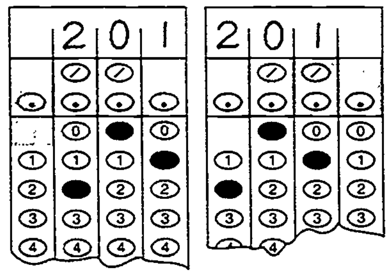
Directions for Student-Produced Response Questions

Each of the remaining 10 questions requires you to solve the problem and enter your answer by marking the ovals in the special grid, as shown in the examples below.

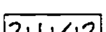
Answer: $\frac{7}{12}$ or 7/12 Answer: 2.5



Answer: 201
Either position is correct.

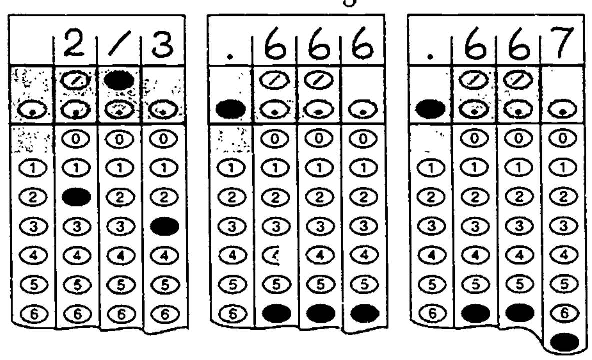


Note: You may start your answers in any column, space permitting. Columns not needed should be left blank.

- Mark no more than one oval in any column.
- Because the answer sheet will be machine-scored, **you will receive credit only if the ovals are filled in correctly.**
- Although not required, it is suggested that you write your answer in the boxes at the top of the columns to help you fill in the ovals accurately.
- Some problems may have more than one correct answer. In such cases, grid only one answer.
- No question has a negative answer.
- **Mixed numbers** such as $2\frac{1}{2}$ must be gridded as 2.5 or 5/2. (If  is gridded, it will be interpreted as $\frac{21}{2}$, not $2\frac{1}{2}$.)

- **Decimal Accuracy:** If you obtain a decimal answer, **enter the most accurate value the grid will accommodate.** For example, if you obtain an answer such as 0.6666 . . . , you should record the result as .666 or .667. **Less accurate values such as .66 or .67 are not acceptable.**

Acceptable ways to grid $\frac{2}{3} = .6666\ldots$



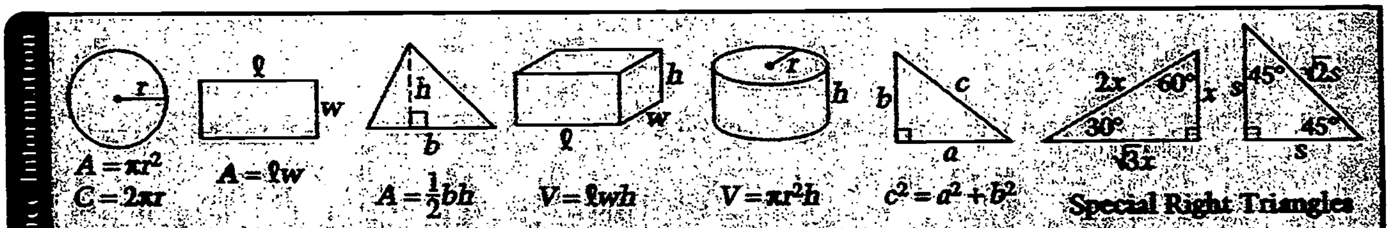
APPENDIX A (cont'd)

Directions for Multiple-Choice Questions with Five Choices

Time—30 Minutes **25 Questions**	**In this section solve each problem, using any available space on the page for scratchwork. Then decide which is the best of the choices given and fill in the corresponding oval on the answer sheet.**

Notes:

1. **The use of a calculator is permitted. All numbers used are real numbers.**
2. **Figures that accompany problems in this test are intended to provide information useful in solving the problems. They are drawn as accurately as possible EXCEPT when it is stated in a specific problem that the figure is not drawn to scale. All figures lie in a plane unless otherwise indicated.**



The number of degrees of arc in a circle is 360.
The measure in degrees of a straight angle is 180.
The sum of the measures in degrees of the angles of a triangle is 180.

Note: This reference information is included in each mathematics section.

1

BEST COPY AVAILABLE

APPENDIX B

Sample Score Report
SPR Items 41-50

Student-Produced Responses					
		Your Answer			
Question Number	Correct Answer(s)	Correct (as gridded)	O Omitted or U Unscorable	Incorrect (as gridded)	Difficulty
41	150	150			M
42	2700	2700			E
43	5		O		M
44	18	18			M
45	1/2 or .5			1/3	M
46	100	100			M
47	25/2 or 12.5	12.5			M
48	36		O		H
49	4/3 < x < 2/1 or 1.33 < x < 2	1.5			H
50	9		O		H
U You made more than one mark in a column; your answer was unscorable.					

E Easy question
M Medium question
H Hard question

From PSAT/NMSQT Score Report for Tuesday, October 11, 1994 Examination

BEST COPY AVAILABLE

18

APPENDIX C

Sample Question with Complete List of All Possible Correct Answers

PSAT/NMSQT
<u>Sample Question</u>

11 If $\frac{k}{10} + \frac{k}{10} + \frac{k}{10} = 3$, what is the value of k ?

<u>Note:</u> This was the first SPR question in Section 4 of the PSAT/NMSQT administered on October 15, 1994.

<u>All Possible Correct Answers</u>

RAW	EDITED *
10	10
010	010
1 0	10
10	10
10.	10.
1 0	10
1 0	10
1 0.	10.
1/.1	1/.1
10	10
10 .	10.
10/1	10/1
10.	10.
10..	10.
10.0	10.0
2/.2	2/.2
20/2	20/2
3/.3	3/.3
30/3	30/3
4/.4	4/.4
40/4	40/4
5/.5	5/.5
50/5	50/5
6/.6	6/.6
60/6	60/6
7/.7	7/.7
70/7	70/7
8/.8	8/.8
80/8	80/8
9/.9	9/.9
90/9	90/9

******** END OF ITEM 41

*See "Editing Rules" section for a discussion of how this column is determined. For example, a raw response of _ _10 is edited to 10 (blank columns are edited out).

APPENDIX D

Additional Sample SPR Items

Sample Item 3

A triangle has a base of length 13 and the other two sides are equal in length. If the lengths of the sides of the triangle are integers, what is the shortest possible length of a side?

This question was on a disclosed SAT I test and had an equated delta of 15.9. Based on a sample size of approximately 3,500 students, 1,100 students omitted the question. Omitting the question was even more popular than answering it correctly. A total of approximately 850 students answered 7, the correct answer. The responses of 1, 13, 6, 6.5, 5, 2, 9, 9.19, 12, 14, 4, 10, 3, 8, 11, 83.5, and 15, listed in order of popularity, were given by at least 20 students. Note that 6.5, 9.19, and 83.5 were given as answers even though the problem stated that the length of the sides of the triangle were integers. It is possible that students answered 83.5 because they thought that the sum of the lengths of the sides of a triangle is 180. (They might have gotten confused because they remembered that the sum of the measures of the angles of a triangle is 180.) However, no high ability students gave 83.5 as a response. The response of 9.19 was not easy to see, but with approximately 50 students giving this response there had to have been some rationale for it. If it is assumed that the triangle is a right triangle, then $x^2 + x^2 = 13^2$ and $x^2 = 84.5$, yielding $x = 9.19$. With the use of calculators, students are trying methods for solving SPR problems which may have been prohibitive without a calculator. They are also giving responses that would never have been considered as answer choices had the same questions been written in the multiple-choice format.

The high ability list showed that approximately 20 high ability students omitted the question, 10 high ability students answered 6, and the number of high ability students giving other responses dropped off from there.

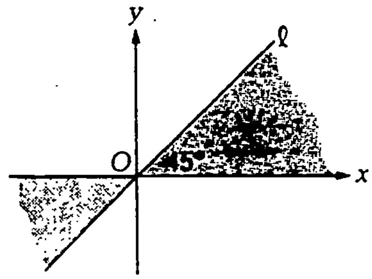
Sample Item 3

Test: SAT I
Equated Delta: 15.9
Sample Size: 3,500*

Response	Number of Students (Approximate)	Number of High Ability Students (Approximate)
Omits	1,100	20
Correct answer 7	850	325
1	300	2
13	175	6
6	150	10
6.5	140	7
5	75	4
2	70	---
9	50	4
9.19	50	8
12	50	1
14	45	2
4	40	1
10	35	7
3	35	1
8	30	2
11	20	3
83.5	20	---
15	20	---

*Note smaller sample size for SAT I tests compared to PSAT/NMSQT tests.

Sample Item 4



Line m (not shown) passes through O in the figure above. If m is distinct from ℓ and the x-axis, and lies in the shaded region, what is a possible slope for m ?

This question was on a PSAT/NMSQT test and had an equated delta of 13.8*. Based on a sample size of approximately 25,000 students, approximately 6,000 students answered this question correctly. Approximately 7,700 students omitted the question. (The high number of omissions was probably due to the printing error.) The responses of 45, 22.5, 1, 2, 0, 30, 90, 3, 35, 135, 25, 40, 3/2, 20, and 22, listed in order of popularity, were given by at least 100 students. The responses of 45, 22.5, 1, and 2 were by far the most popular wrong answers with 2,500 students answering 45, 1,800 answering 22.5, 1,500 answering 1, and 1,100 answering 2. This is a question with multiple correct answers in a range. Any number between 0 and 1 that can be gridded on the grid is a correct answer to this question. The most popular correct answer was 1/2 (answered by 4,000 students) and the next most popular correct answer was 1/3 (answered by only 600 students). The responses of 45, 22.5, 30, 90, 35, 135, 25, 40, 20, and 22 were all given by students probably thinking of angle measures. These types of responses would never have been considered as answer choices if this question were written in a multiple-choice format.

The high ability list showed that of the high ability students, approximately 300 students omitted the question; 300 students answered 2; 200 students answered 1; and 100 students answered 22.5. The number of high ability students giving other responses dropped off from there.

*This item was dropped from scoring on the 1993 PSAT/NMSQT because of a printing error in some of the test books.

Sample Item 4

Test: PSAT/NMSQT
Equated Delta: 13.8
Sample Size: 25,000

Response	Number of Students (Approximate)	Number of High Ability Students (Approximate)
Omits	7,700	300
Correct answer $0 < x < 1$	6,000	2,900
45	2,500	30
22.5	1,800	100
1	1,500	200
2	1,100	300
0	550	50
30	300	50
90	300	1
3	250	30
35	200	20
135	200	2
25	150	20
40	150	30
3/2	150	30
20	150	20
22	100	10

23

Sample Item 5

> Set S consists of all multiples of 3 between 10 and 25. Set T consists of all multiples of 4 between 10 and 25. What is one possible number that is in set S but <u>not</u> in set T?

This question was on a PSAT/NMSQT test and had an equated delta of 9.2. Based on a sample size of approximately 25,000 students, the most popular response, after the correct answers, was to omit the question. In this sample, approximately 1,800 omitted the question. This was a question with multiple correct answers. For this question, however, there are only three correct answers as opposed to sample question 4 that had multiple correct answers in a range. Responses of 15, 18, and 21 are all correct. There were 14,000 students who answered 15; 2,000 who answered 18; and 2,000 who answered 21. The most popular correct answer by far was 15. Listed in order of popularity, the most popular incorrect answers were 16, 3, 13, 7, 20, 9, 12, 6, 8, 5, and 25 given by at least 100 students. The numbers 16 and 20 are in set *T* but not in set *S*, suggesting that students answered the question in reverse. These two responses would probably have appeared as answer choices in a multiple-choice version of this question. Although 12 may also have appeared as an answer choice, responses of 3, 13, 7, 9, 6, 8, 5, and 25 would probably not have been considered.

The high ability list showed that the response of 16 was given by 80 high ability students. The number of high ability students who gave any of the other responses was quite a bit less than 80.

Sample Item 5

Test: PSAT/NMSQT
Equated Delta: 9.2
Sample Size: 25,000

Response	Number of Students (Approximate)	Number of High Ability Students (Approximate)
Correct answer 15, 18, 21	18,000	4,300
omit	1,800	10
16	1,000	80
3	600	15
13	400	10
7	400	8
20	350	30
9	300	15
12	200	15
6	150	10
8	150	4
5	150	---
25	130	2

Sample Item 6

For all nonnegative numbers a, let $\boxed{a}$ be defined by $\boxed{a} = \frac{\sqrt{a}}{3}$. If $\boxed{a} = 2$, what is the value of a ?

This question was on a disclosed SAT I test and had an equated delta of 13.4. Out of the 3,500 sample, approximately 1,500 students gave the correct answer of 36. There were 600 students who omitted the question. The next most popular responses, listed in order of popularity, were .471, 6, .47, 2, 2.45, 2/3, 4, 1.41, 3, 2.44, and 12 given by at least 20 students. Responses of .471, .47, 2.45, 1.41, and 2.44 would not have been included as answer choices if this question was written in a multiple-choice format because they are obtained by using a calculator. The PSAT/NMSQT and SAT I tests only permit the use of a calculator, they do not require its use. The responses of .471 and .47 are approximately equal to $\frac{\sqrt{2}}{3}$ which is $\boxed{2}$. To actually solve this problem, substitute 2 for $\boxed{a}$ so that $2 = \frac{\sqrt{a}}{3}$ and $6 = \sqrt{a}$. The correct answer is 6^2 which equals 36. Any student who took the square root of 6 instead of squaring 6 in this last step, would obtain an answer of 2.45 or 2.44, two of the other popular responses. The response of 1.41 is approximately equal to $\sqrt{2}$.

The high ability list showed that 21 high ability students answered .471. The number of high ability students who gave other responses was 8 or less.

Sample Item 6

Test: SAT I
Equated Delta: 13.4
Sample Size: 3,500*

Response	Number of Students (Approximate)	Number of High Ability Students (Approximate)
Correct answer 36	1,500	300
omits	600	1
.471	350	21
6	125	2
.47	100	1
2	100	1
2.45	100	8
2/3	70	---
4	50	---
1.41	30	---
3	25	---
2.44	20	1
12	20	---

*Note smaller sample size for SAT I tests compared to PSAT/NMSQT tests.

Sample Item 7

> In a stack of six cards, each card is labeled with a different integer 0 through 5. If two cards are selected at random without replacement, what is the probability that their sum will be 3 ?

This question was on a disclosed SAT I test and had an equated delta of 18.2, making it the most difficult SPR question discussed in this paper. Based on a sample size of approximately 3,500, approximately 900 students omitted the question. A response of 1/3, or its equivalent, was the next most popular answer with 500 students gridding this answer. There were 400 students who gave the correct answer of 2/15 or its equivalent. Listed in order of popularity, other responses given by at least 20 students were 2/5, 1/6, 2/3, 2, 1/5, 1/2, 1/15, 1/9, 1/18, 3/5, 1/4, 1/30, 1, 1/10, 2/25, 1/12, and 3.

The fact that there were many answers obtained by 20 students or more, indicates that this was a good question to ask in the SPR format. A 5-choice question would not have been able to capture all of the misconceptions students may have on probability. Also, in the 5-choice format, a student would see an answer choice with 15 in the denominator (the correct answer) and be given a big hint for solving the problem.

The high ability list showed that 30 high ability students omitted the question and 30 high ability students answered 1/15. A response of 1/3 was given by 20 high ability students and the number of high ability students giving other responses dropped off from there.

Sample Item 7

Test: SAT I
Equated Delta: 18.2
Sample Size: 3,500*

Response	Number of Students (Approximate)	Number of High Ability Students (Approximate)
Omits	900	30
1/3	500	20
Correct answer 2/15	400	200
2/5	160	8
1/6	160	7
2/3	140	1
2	120	---
1/5	120	9
1/2	100	1
1/15	60	30
1/9	40	10
1/18	40	3
3/5	40	1
1/4	40	2
1/30	30	14
1	30	---
1/10	25	6
2/25	25	2
1/12	20	6
3	20	---

*Note smaller sample size for SAT I tests compared to PSAT/NMSQT tests.

Sample Item 8

In a certain factory, 0.2 percent of a batch of microchips are defective. If this batch contains 4 defective microchips, how many microchips are in the batch?

This question was on a PSAT/NMSQT test and had an equated delta of 13.0. Based on a sample size of approximately 25,000, there were 8,200 students who gave a correct response of 2000. There were 5,400 students who answered 20 and 4,000 students who omitted the question. Listed in order of popularity, other incorrect answers given by at least 100 students were 200, 80, .8, 8, 500, 800, 400, .008, 16, 5, .05, 1000, 100, 50, 2, and 125. It is interesting to note how popular responses that involved an incorrect use of 0.2 percent were in this sample. Responses of 20 and 200 are obtained by incorrectly working with 0.2 percent. If students use .2 or .02 instead of .002, they arrive at answers of 20 or 200, respectively. Given that students are permitted to use a calculator when taking this test, it is surprising that the number of students giving these responses is so high. Even the number of high ability students answering 20 or 200 is high. Responses of 80, .8, 8, and .008 are obtained by using 20, .2, 2, and .002, respectively, and multiplying rather than dividing by these numbers to obtain an answer. Clearly, having a calculator does not guarantee that students will obtain correct answers to questions on the tests.

The high ability list showed that 700 high ability students in the sample answered 20, and 170 high ability students answered 200. There were 80 high ability students who omitted the question and the number of high ability students who gave other responses dropped off from there.

Sample Item 8

Test: PSAT/NMSQT
Equated Delta: 13.0
Sample Size: 25,000

Response	Number of Students (Approximate)	Number of High Ability Students (Approximate)
Correct answer 2,000	8,200	3,400
20	5,400	700
Omits	4,000	80
200	1,400	170
80	800	17
.8	800	4
8	450	---
500	200	12
800	200	6
400	200	4
.008	150	3
16	150	1
5	150	1
.05	140	---
1,000	120	4
100	120	2
50	120	---
2	100	4
125	100	1