The College Board

Research Notes

Office of Research and Development RN-14, December 2001

Ensuring Comparable Scores on the SAT® I: Reasoning Test

Ida M. Lawrence and Amy Elizabeth Schmidt

INTRODUCTION

The SAT® I: Reasoning Test is administered seven times a year. Primarily for security purposes, several different test forms are given at each administration. How is it possible to compare scores obtained from different test forms and from different test administrations? The purpose of this paper is to provide an overview of the statistical procedures used to produce comparable scores on different forms of the SAT I.

Scores on the SAT I are reported on a scale that ranges from 200 to 800, in increments of 10. Each of these scaled scores represents a particular level of achievement in either verbal or mathematical reasoning as measured by the SAT I. In order to obtain comparable scaled scores across different forms of the test, several steps must be taken. First, SAT I formula scores are calculated based on each student's number of correct and incorrect answers. Then, different forms are equated to produce comparable formula scores across forms. Formula scores are then converted to scaled scores, and finally, interpretive information is provided. Each of these concepts is dealt with in more detail below.

CALCULATING FORMULA SCORES

Before students' scores can be placed on the 200 to 800 scale, formula scores must first be computed. Formula scores are used to adjust the scores for guessing. The following rules are used to compute formula scores:

KEYWORDS:

SAT I
Equating
Scaling

- Each correct answer is worth one point.
- A fraction of a point is subtracted for each incorrect answer to a multiple-choice question. One-fourth of a point is subtracted for each incorrect answer to multiple-choice questions with five answer options. One-third of a point is subtracted for each incorrect answer to multiple-choice questions with four answer options. No points are subtracted for incorrect answers to student-produced response questions.
- No points are added or subtracted for omitted questions.

For example, in the verbal section of the SAT I, which consists of 78 items, each having five options, the following equation is used to calculate the formula score:

$$FS = C - kW$$

Where:

FS = Formula score
C = Total number of items correct
W = Total number of items incorrect
k = $1/(n-1)$
n = Number of options in a multiple choice item (k = .25 for verbal items)

Calculating formula scores for the math section is slightly more complicated because of the three distinct types of items. In the 60-item math section, there are 35 five-option multiple-choice items, 15 four-option quantitative comparison items, and 10 student-produced response items, or "grid-ins." Therefore, the formula score equation for the math section is:

$$FS = C - k_1W_1 - k_2W_2$$

Where:

FS = Formula score
C = Total number of items correct
W_1 = Total number of five-option multiple-choice items incorrect

W_2= Total number of four-option quantitative comparison items incorrect

k_1 = $1/(n_1 - 1)$

k_2 = $1/(n_2 - 1)$

n_1 = no. of options in a multiple-choice item (k_1 = .25 for five-option items)

n_2 = no. of options in a quantitative comparison item (k_2 = .3333 for four-option items)

A student who takes an easier form of the SAT I will receive a higher formula score compared to the score he or she would obtain on a more difficult form of the SAT I. However, equating procedures are then used to produce scaled scores that adjust for differences in test difficulty.

SCORE EQUATING AND SCALING

Detailed content and statistical specifications are used to assemble each new form of the SAT I. One goal of the test assembly process is to make all test forms equivalent in difficulty. In practice, it is not possible to produce test forms that are exactly equivalent in difficulty, and a statistical procedure, referred to as score equating, is used to ensure that scores on different forms of the SAT I are comparable. Thus, the purpose of equating is to adjust scores for minor differences in test difficulty from form to form, so that a score represents the same level of achievement regardless of difficulty of a particular form.

In order to accomplish this equating most efficiently, it is necessary that the various forms of the SAT I be linked in some way. The procedures used to equate forms of the SAT I are based on a data collection plan that links each new form to several previously scaled forms. Each new form is administered with several "anchor tests," each of which have been administered with a previous form at a previous administration. An anchor test is a miniature version of the SAT I, and because it appears in both the new form that needs to be equated as well as in previous forms, the anchor test provides a link between the current form and the previous forms. The variable section of the SAT I, which is used to administer the anchor test, looks similar to a section in the verbal or math test. However, the responses to the questions in the variable section do not count toward the total score. In terms of content and statistical properties, the anchor test is a miniature version of the total test.

Separate equating analyses are carried out for the verbal scores and the math scores. Equating analyses are based on paired samples of students who took either the new form and a particular anchor test or a previous form and the same anchor test. The samples are representative of the student population. Equating a new form of the SAT I to a previous form of the SAT I involves three basic steps.

- First, formula scores on the anchor test that was administered to both samples of test-takers are used to adjust for differences in the ability levels of the students who took the new form and those who took the previous form.
- Second, a variety of mathematical models are employed to develop equating functions relating formula scores on the new form and formula scores on the previous form. Equating methods based on linear models, equipercentile models, and item response theory are used. Equating formula scores on the new and previous form involves an evaluation of the relative difficulty of the two forms after adjusting for differences in ability of the samples that took the previous form and the new form.
- Third, equated formula scores on the SAT I are converted to scaled scores ranging from 200 to 800 in increments of 10 points. Converting raw scores into scaled scores is a mathematical process that results in an equation or table of values that relates each formula score on the new form to a corresponding score on the reporting scale. This is possible because the scaled score corresponding to each raw score on the previous form is known.

After these three steps have been completed for each of several links to previous forms, equating results from each link are combined together to determine the table for converting formula scores on the new form to scaled scores. Every new form has a unique table for converting equated formula scores to scaled scores.

As a result of an equating analysis, scaled scores on different forms of the SAT I will vary as a function of differences in the difficulty of the forms.

For instance, a formula score of 43 on a relatively easy form might correspond to a scaled score of 540. On a more difficult form, the same formula score of 43 might correspond to a scaled score of 560. However, the equating analysis ensures that the 540 and 560 are on the same SAT scale, and can be directly compared with one another.

Another result of equating is that differences in formula scores may result in differences in magnitude of conversion to scaled scores within the same form, depending on where the student falls on the scale. An example involving the verbal portion of a particular SAT I form is provided to illustrate this point; keep in mind, however, that these results are different across forms and may not generalize to other forms. For this example, three hypothetical students were used: a low scoring student, or a student who got 21 correct and 57 incorrect on the verbal test; a mid-range scoring student, or a student who got 44 correct and 34 incorrect; and a high scoring student, or a student who got 68 correct and 10 incorrect. Note that in this example, it is assumed that the students completed the entire verbal form and got each item either correct or incorrect. Formula scores were calculated from C (the total number correct) and W (the total number incorrect; FS 1 in Table 1), and these were converted to scaled scores using the conversion table that was constructed for this form. To examine what would happen if the number of items correct increased, one point was added to each raw score, and the associated formula score was calculated, along with its scaled score (FS 2 in Table 1). For illustrative purposes, this was repeated two more times (FS 3 and FS 4 in Table 1).

As can be seen in this example, an increase in one correct item leads to a scaled score change of 10 points for low and high scoring students, but no change for mid-range scoring students. An increase of two correct items results in a scaled score change of 20 points for low and high scoring students, but only 10 for mid-range scoring students, while an increase of three correct items results in a scaled score change of 30 points for low scoring students, 20 points for mid-range students, and 40 points for high scoring students.

This example, however, may not be a terribly good reflection of reality, since students complete the SAT at different rates. The average completion rate for the test form was approximately 84 percent; therefore, the example was modified in Table 2 to reflect varying completion rates. It was assumed that lower scoring students would complete at a lower rate, and therefore a 68 percent completion rate was assumed for low scoring students, 84 percent assumed for mid-range students, and 100 percent for high scoring students.

Again, increases in the scaled score vary depending on where the student is on the scale. An increase in one correct item leads to a scaled score change of 10 points for low and high scoring students, but no change for mid-range scoring students. An increase of two correct items results in a scaled score change of 20 points for low scoring students, 10 for mid-range scoring students, and 30 points for high scoring students, while an increase of three correct items results in a scaled score change of 30 points for low scoring students, 10 points for mid-range scoring students, and 40 points for high scoring students.

TABLE 1
CHANGES IN FORMULA SCORE TO SCALED SCORE CONVERSIONS AS A FUNCTION OF STUDENT PERFORMANCE: ALL ITEMS COMPLETED

	FS 1	SS	FS 2	SS	FS 3	SS	FS 4	SS
Low	7 (21C, 57W)	310	8 (22C, 56W)	320	9 (23C, 55W)	330	11 (24C, 54W)	340
Mid	36 (44C, 34W)	500	37 (45C, 33W)	500	38 (46C, 32W)	510	39 (47C, 31W)	520
High	66 (68C, 10W)	700	67 (69C, 9W)	710	68 (70C, 8W)	730	69 (71C, 7W)	740

SS = scaled score, FS = formula score, C = number of items correct, W = number of items incorrect.

TABLE 2
CHANGES IN FORMULA SCORE TO SCALED SCORE CONVERSIONS AS A FUNCTION OF STUDENT PERFORMANCE: VARIABLE NUMBER OF ITEMS COMPLETED

	FS 1	SS	FS 2	SS	FS 3	SS	FS 4	SS
Low (68% Comp)	13 (21C, 32W, 25O)	360	14 (22C, 31W, 25O)	370	16 (23C, 30W, 25O)	380	17 (24C, 29W, 25O)	390
Mid (84% Comp)	39 (44C, 22W, 12O)	520	40 (45C, 21W, 12O)	520	41 (46C, 20W, 12O)	530	42 (47C, 19W, 12O)	530
High (100% Comp)	66 (68C, 10W, 0O)	700	67 (69C, 9W, 0O)	710	68 (70C, 8W, 0O)	730	69 (71C, 7W, 0O)	740

SS = scaled score, FS = formula score, C = number of items correct, W = number of items incorrect, O = number of items omitted.

INTERPRETING SCALED SCORES

When an organization provides test score information, it also has a professional obligation to provide interpretive information as well (APA, AERA, and NCME, 1999). This information can take various forms, but a common type of information that is provided is referred to as norms. In very broad terms, norms can be thought of as any information that provides a frame of reference for interpreting test scores. In this light, the SAT program provides a wealth of normative information to test-takers and users of test scores on an annual basis. An annual report on college-bound seniors, released by the College Board each August, presents information on national and state SAT score means, as well as mean scores by gender, ethnic/racial background, parental education, high school courses taken, and other variables of interest (e.g., College Board, 2001). In addition, the *College Board Program Handbook* contains normative information about the scaled scores for a particular testing year, presented in the form of percentile ranks based on college-bound seniors who took the SAT I and graduated from high school in the previous year (College Board, 2001a). This publication also provides information on the effects of repeat testing and coaching; statistical test characteristics, such as reliability, difficulty level and completion rates; and validity information, all of which are important for score interpretation. Information found in printed publications is also available on the College Board Web site: www.collegeboard.com.

In addition to the information provided by College Board publications and Web site, which are mostly used by test score users and policymakers, the College Board provides important interpretive information to test-takers, primarily in the form of percentile ranks. Each scaled score has an associated percentile rank, which is the percentage of test-takers that scored below that particular scaled score. Both national and state level percentile ranks are provided to the test-takers so that they can compare themselves to students across the nation and within their own state. Although the percentile ranks associated with each scaled score may change from year to year, the meaning of scaled scores stays the same across years.

SUMMARY

The methods used to equate the various forms of the SAT I in order to put them on a common scale were developed in keeping with the highest psychometric and technical standards (APA, AERA, and NCME, 1999; Donlon, 1984; Kolen and Brennan, 1995). These methods assure that test users can be confident that the meaning of SAT I scores will remain the same from year to year, regardless of the variation in difficulty across test forms or the variation in ability across test administrations.

The authors are Amy Elizabeth Schmidt, director of higher education and evaluation research at the College Board, and Ida M. Lawrence, executive director, School and College Services Division, at Educational Testing Service.

REFERENCES

American Psychological Association, American Educational Research Association, and National Council on Measurement in Education (1999). *Standards for educational and psychological testing.* Washington, DC: Author.

College Board (2001). *College-bound seniors: A profile of SAT program test-takers.* New York, NY: College Entrance Examination Board.

College Board (2001a). *College Board Program Handbook.* New York, NY: College Entrance Examination Board.

Donlon, T. F. (Ed.) (1984). *The College Board technical handbook for the scholastic aptitude test and achievement tests.* New York, NY: College Entrance Examination Board.

Kolen, M. J., & Brennan, R.L. (1995). *Test equating: Methods and practices.* New York, NY: Springer-Verlag.

**For more information or additional copies of this report, please write to:
Office of Research, The College Board, 45 Columbus Avenue, New York, NY 10023-6992,
or visit the Research Library on our Web site at: www.collegeboard.com.**

12/01