

# A Historical Perspective on the SAT<sup>®</sup> 1926–2001

Ida Lawrence, Gretchen W. Rigol, Thomas Van Essen,  
and Carol A. Jackson



# A Historical Perspective on the SAT<sup>®</sup> 1926–2001

Ida Lawrence, Gretchen W. Rigol, Thomas Van Essen,  
and Carol A. Jackson

Ida Lawrence is Executive Director, Division of School and College Services at Educational Testing Service (ETS).

Gretchen W. Rigol is Vice President, International and Special Services at the College Board.

Thomas Van Essen is Director, SAT® and PSAT/NMSQT® Programs, Division of School and College Services at ETS.

Carol A. Jackson is Assessment Specialist in Mathematics at ETS.

Researchers are encouraged to freely express their professional judgment. Therefore, points of view or opinions stated in College Board Reports do not necessarily represent official College Board position or policy.

#### *The College Board: Expanding College Opportunity*

The College Board is a national nonprofit membership association whose mission is to prepare, inspire, and connect students to college and opportunity. Founded in 1900, the association is composed of more than 4,200 schools, colleges, universities, and other educational organizations. Each year, the College Board serves over three million students and their parents, 22,000 high schools, and 3,500 colleges through major programs and services in college admissions, guidance, assessment, financial aid, enrollment, and teaching and learning. Among its best-known programs are the SAT®, the PSAT/NMSQT®, and the Advanced Placement Program® (AP®). The College Board is committed to the principles of excellence and equity, and that commitment is embodied in all of its programs, services, activities, and concerns.

For further information, visit [www.collegeboard.com](http://www.collegeboard.com).

Additional copies of this report (item #994738) may be obtained from College Board Publications, Box 886, New York, NY 10101-0886, 800 323-7155. The price is \$15. Please include \$4 for postage and handling.

Copyright © 2002 by College Entrance Examination Board. All rights reserved. College Board, Advanced Placement Program, AP, SAT, and the acorn logo are registered trademarks of the College Entrance Examination Board. PSAT/NMSQT is a registered trademark jointly owned by both the College Entrance Examination Board and the National Merit Scholarship Corporation. Other products and services may be trademarks of their respective owners. Visit College Board on the Web: [www.collegeboard.com](http://www.collegeboard.com).

Printed in the United States of America.

## *Contents*

I. Introduction .....	1
II. Early Versions of the SAT (1926–1930) .....	1
III. Changes to the Verbal Portion of the SAT Since 1930 .....	2
IV. Changes to the Mathematical Portion of the SAT Since 1930 .....	6
Data Sufficiency Item .....	7
Quantitative Comparison Item .....	7
Student-Produced Response Questions .....	9
V. What the SAT Measures Today .....	12
VI. Conclusion .....	12
References .....	13
<i>Tables</i>	
1. Percentage of SAT Test-Takers by Racial/Ethnic Background 1973–2001 .....	1
2. Changes to Format of SAT-Verbal 1958–2001 .....	6
3. Changes to Format of SAT-Math 1942–2001 .....	8

# I. Introduction

The current debate over admissions test requirements at the University of California has sparked an interest in what is measured by the various tests—in particular, what is measured by the SAT®. This paper uses a historical perspective to present an overview of changes in the content of the College Board’s SAT I: Reasoning Test (popularly referred to as the SAT).

The SAT has been reconfigured several times over the years. Each redesign was intended to make the test more useful to students, teachers, high school counselors, and college admissions staff. Since 1970 test modifications have focused on the following goals:

- Ensuring that test content is balanced and appropriate for test-takers with widely different cultural and educational backgrounds
- Ensuring that test performance is reliably measured and that the test effectively differentiates among test-takers across the full range of scores
- Reducing the influence of speed on test performance
- Reducing the effects of special preparation on test performance
- Ensuring that test content is consistent with changes in school-based learning
- Ensuring adequate comparability of scores over extended periods of time

Regarding the last point, it should be noted that most changes have been intentionally gradual, in part to ensure that scores from new versions of the test could be equated to earlier versions.

Although this paper does not discuss changes in the test-taking population, it is important to keep in mind that the profile of those who take the SAT has altered

dramatically since its beginnings 75 years ago. About 8,000 young men took what was then called the “Scholastic Aptitude Test” at its first administration in 1926, but more than 2.7 million SAT tests were administered to young men and women in the United States and abroad in the 2000–2001 testing year. The steepest increases in test volume since 1973 have been among students of Asian or Hispanic/Latino descent; the proportion of African American test-takers has also increased. The proportion of white test-takers decreased from 87 percent in 1973 to 66 percent in 2001. For more information on the racial/ethnic background of present-day test-takers, see Table 1.

# II. Early Versions of the SAT® (1926–1930)

The 1926 version of the SAT bears little resemblance to the current test. It contained nine subtests: seven with verbal content (definitions, classification, artificial language, antonyms, analogies, logical inference, and paragraph reading) and two with mathematical content (number series and arithmetical problems). The time limits were quite stringent: 315 questions were administered in 97 minutes. Early versions of the SAT were quite speeded—as late as 1943, students were told that they should not expect to finish. Even so, many of the early modifications to the test were aimed at providing more liberal time limits. In 1928 the test was reduced to seven subtests administered in 115 minutes, and in 1929, to six subtests.

In addition to seeking appropriate time limits, developers of these early versions of the SAT were also concerned with the possibility that the test would influence educational practices in negative ways. On the basis of empirical research that looked at the effects of pretest

TABLE 1

**Percentage of SAT Test-Takers by Racial/Ethnic Background 1973–2001**

	1973	1978	1983	1988	1993	1998	2001
American Indian or Alaskan Native		<1%	<1%	<1%	<1%	<1%	<1%
Asian, Asian American, or Pacific Islander	2	3	4	6	8	9	9
African American or Black	7	9	9	9	11	11	11
Hispanic or Latino Background	1	3	3	5	7	8	9
White	87	83	81	77	70	67	66
Other	1	2	2	1	2	3	4
Number Responding	784,848	893,767	875,475	1,055,557	953,119	1,049,773	1,072,577

practice on the various question types, antonyms and analogies were used because research indicated they were less responsive to practice than were some of the other question types (Coffman, 1962).

Beginning in 1930 the SAT was split into two sections, one portion designed to measure “verbal aptitude” and the other to measure “mathematical aptitude.” Reporting separate verbal and mathematical scores allowed admissions staff to weight the scores differently depending on the type of college and the nature of the college curriculum.

### III. Changes to the Verbal Portion of the SAT Since 1930

Verbal tests administered between 1930 and 1935 contained only antonyms, double definitions, and paragraph reading. In 1936 analogies were again added. Verbal tests administered between 1936 and 1946 included various combinations of antonyms (100 questions), analogies (50 questions), double definitions (50 questions), and paragraph reading (50 questions). The amount of time to complete these tests ranged between 80 and 115 minutes, depending on the year the test was taken.

The antonym question type in use between 1926 and 1951 was called the “six-choice antonym.” Test-takers were given a group of four words and told to select the two that were “opposite in meaning” (according to the directions given in 1934) or “most nearly opposite” (according to the 1943 directions). These were called “six-choice” questions because there were six possible pairs of numbers from which to choose: (1, 2), (1, 3), (1, 4), (2, 3), (2, 4), and (3, 4).

Here is an example of medium difficulty from 1934:

gregarious<sub>1</sub> solitary<sub>2</sub> elderly<sub>3</sub> blowy<sub>4</sub>

(Answer: 1, 2)

Here is a difficult example from 1943:

1-divulged 2-esoteric 3-eucharistic 4-refined

(Answer: 1, 2)

In the 1934 edition of the test, test-takers were asked to do 100 of these questions in 25 minutes. They were given no advice about guessing strategies, and the

instructions had a quality of inscrutable moralism: “Work steadily but do not press too hard for speed. Accuracy counts as well as speed. Do not penalize yourself by careless mistakes.”

In 1943 test-takers were given an additional 5 minutes to complete 100 questions, but this seeming generosity was compensated for by a set of instructions that seem bizarre by today’s standards: “Work steadily and as quickly as is consistent with accuracy. The time allowed for each subtest has been fixed so that very few test-takers can finish it. Do not worry if you cannot finish all the questions in each subtest before time is called.”

In 1952 the antonym format was changed to the more familiar five-choice question. Here is a sample from 1960:

VIRTUE: (A) regret (B) hatred (C) penalty  
(D) denial (E) depravity

(Answer: E)

The five-choice question is a more direct measure of vocabulary knowledge than the six-choice question, which is more like a puzzle. There are two basic ways to solve the six-choice antonym. The first is to read the four words, grasp them as a whole, and determine which two are opposites. This approach requires the ability to keep a large chunk of material in the clipboard of short-term memory while manipulating it and comparing it to the resources of vocabulary knowledge that one brings to the testing situation. The other approach is to apply a simple algorithm to the problem: “Is the first word the opposite of the second word? If not, is the first word the opposite of the third word? If not, is the first word...” and so forth until all six choices have been evaluated.

Most test-takers probably used some combination of the two methods, first trying the holistic approach and if that didn’t work, using the more systematic approach. The latter approach probably took longer than the former; given the tight time constraints of the test at this time (18 seconds an item!), test-takers who relied solely on the systematic approach were at a disadvantage.

Note that in one of the samples above (1-divulged 2-esoteric 3-eucharistic 4-refined), the vocabulary is quite specialized by the standards of today’s test. The word “eucharistic” would never be used today, because it is a piece of specialized vocabulary that is more familiar to Roman Catholics and other Christians than to non-Christians. Even the sense of “divulged” as the opposite of “esoteric” is obscure, with “divulged” taking the sense of “revealed” or “given out,” while “esoteric” has the sense of “secret” or “designed for, or appropriate to, an inner circle of advanced or privileged disciples.”

---

The double definition question type was a precursor of the sentence completion question that served as a complement to antonyms by focusing on vocabulary knowledge from another angle. This question type was used from 1928 to 1941.

Here is an example of medium difficulty from 1934:

A \_\_\_\_ is a venerable leader ruling by \_\_\_\_ right.  
mayor<sub>1</sub> patriarch<sub>2</sub> minister<sub>3</sub> general<sub>4</sub>  
paternal<sub>1</sub> military<sub>2</sub> ceremonial<sub>3</sub> electoral<sub>4</sub>

(Answer: 2, 1)

This is a fairly straightforward measure of vocabulary knowledge, although it too contains an element of puzzle solving as the test-taker is required to choose among the 16 possible answer choices. In 1934 test-takers were given 50 of these questions to answer in 20 minutes.

A question type called paragraph reading was featured on the test from 1926–1945. These questions presented test-takers with one or two sentences of 30–70 words and asked them to identify the word in the paragraph that needed to be changed because it spoiled the “sense or meaning of the paragraph as a whole.” From 1926 through 1938 test-takers were asked to cross out the inappropriate word, and from 1939 through 1946 they were asked to choose from one of 7 to 15 (depending on the year) numbered words.

Here is an easy example from 1943:

Everybody<sub>1</sub> in college who knew<sub>2</sub> them at all was convinced<sub>3</sub> to see what would come<sub>4</sub> of a friendship<sub>6</sub> between two persons so opposite<sub>7</sub> in tastes, and appearances.

(Answer: 3)

The task here is less a reasoning task than a proofreading task, and the only real source of difficulty is the similarity in sounds between “convinced” and “curious.” A careless test-taker might be unable to see “convinced” as the problem because he or she simply corrected it to “curious.”

Here is a difficult (in more senses than one) example from the same year:

At last William bade his knights draw off<sub>1</sub> for a space<sub>2</sub>, and bade the archers only continue the combat. He feared<sub>3</sub> that the English, who had no<sub>4</sub> bowmen on their side, would find the rain of arrows so unsupportable<sub>5</sub> that they would at last break their line and charge<sub>6</sub>, to drive off their tormentors<sub>7</sub>.

(Answer: 3)

This question tests reading skills, but it also tests informal logic and reasoning. The key to the difficulty is that as the test-taker reads the beginning of the second sentence, he or she probably assumes that William is English—it is only when the reader figures out that the English have no bowmen that he or she realizes that William must be fighting the English. Here the issue of outside knowledge comes in. Readers who are familiar with English history know that a William who used archers successfully was William the Conqueror in his battles against the English. This knowledge imparts a terrific advantage, especially given the time pressure. It also helps if the test-taker knows enough about military matters to accept the idea that a military leader might want the opposing forces to charge.

The paragraph-reading question was dropped after 1945. The verbal test that appeared in 1946 contained antonyms, analogies, sentence completions, and reading comprehension. With the exception of antonyms, this configuration is similar to that of today’s SAT and represents a real break with the test that existed before. Changes were made in the interest of making the test more relevant to the process of reading: the test is still a verbal reasoning test, but the balance has shifted somewhat from *reasoning* to *verbal*.

Critics of the SAT often point to its heritage in the intelligence tests of the early years of the last century and condemn the test on account of its pedigree, but it is worth noting that by 1946 those question types that were most firmly rooted in the traditions of intelligence testing had fallen by the wayside, replaced by questions that were more closely allied to English and language arts. “The double definition is a relatively restricted form; the sentence completion permits one the use of a much broader range of material. In the sentence completion item the candidate is asked to do a kind of thing which he does naturally when reading: to make use of the element of redundancy inherent in much verbal communication to obtain meaning from something less than the complete communication” (Loret, 1960, p. 4). The change to reading comprehension items was made for a similar reason. “The paragraph reading item probably tends to be esoteric, coachable, and relatively inefficient, while the straightforward reading comprehension is commonplace, probably non-coachable, and reasonably efficient in that a number of questions are drawn from each passage” (Loret, 1960, pp. 4–5).

This shift in emphasis is seen most clearly by comparing the paragraph reading questions discussed above with the reading comprehension questions that replaced them. By the 1950s about half of the testing time in the verbal section was devoted to reading. Here is a short (at this time the passages ranged between 120 words and 500 words) reading comprehension passage that

appeared in the descriptive booklet made available to students in 1957:

Talking with a young man about success and a career, Doctor Samuel Johnson advised the youth “to know something about everything and everything about something.” The advice was good—in Doctor Johnson’s day, when London was like an isolated village and it took a week to get the news from Paris, Rome, or Berlin. Today, if a man were to take all knowledge for his province and try to know something about everything, the allotment of time would give one minute to each subject, and soon the youth would flit from topic to topic as a butterfly from flower to flower and life would be as evanescent as the butterfly that lives for the present honey and moment. Today commercial, literary, or inventive success means concentration.

The questions that followed were mostly what the descriptive booklet described as “plain sense” questions. Here is an easy to medium-difficult example:

According to the passage, if we tried now to follow Doctor Johnson’s advice, we would

- (A) lead a more worthwhile life
- (B) have a slower-paced, more peaceful, and more productive life
- (C) fail in our attempts
- (D) hasten the progress of civilization
- (E) perceive a deeper reality

(Answer: C)

Although this question can be answered without making any complicated inferences, it does ask the test-taker to make a connection between the text and his or her own life.

Here is a question in which test-takers were asked to evaluate and pass judgment on the passage:

In which one of the following comparisons made by the author is the parallelism of the elements least satisfactory?

- (A) Topics and flowers
- (B) The youth and the butterfly
- (C) London and an isolated village
- (D) Knowledge and province
- (E) Life and the butterfly

(Answer: E)

Here the test writers were essentially asking test-takers to identify a serious flaw in the logic and composition of the passage. According to the rationale provided in the descriptive book, “the comparison” made in (E) “is a little shaky. What the author really means is that human life would be like the life of a butterfly—aimless and evanescent—not that human life would be like the butterfly itself. The least satisfactory comparison, then, is E.” This question attempts to measure a higher-order critical-thinking skill.

Verbal tests administered between 1946 and 1957 typically contained between 107 and 170 questions. Testing time ranged between 90 and 100 minutes. Beginning in 1958 test length was reduced to 90 questions given in 75 minutes; this format remained unchanged until 1974, when the Test of Standard Written English was added and the verbal and math tests were reduced from 75 to 60 minutes. To accommodate the shorter testing time and still administer a sufficient number of questions to maintain test reliability, the mix of discrete and passage-based questions had to be altered. Another change in test format took place in 1978 when the mix of medium-length and long passages was adjusted. This change was also made to control test speededness.

The next major changes to the verbal test took place in 1994, when antonyms were removed. The rationale for making this change was that antonym questions present words without a context and encourage rote memorization. Another important change was an increase in the percentage of questions associated with passage-based reading material. For SAT tests administered between 1974 and 1994, the percentage of passage-based reading questions was 29 percent. To send a signal to schools about the importance of reading, in 1994 passage-based reading questions were increased to 50 percent. This added reading necessitated an increase in testing time and a decrease in the total number of questions. Major changes to the verbal test introduced in 1994 were as follows (Curley and May, 1991):

- Emphasis on critical reading and reasoning skills
- Reading material that is accessible and engaging
- Passages ranging in length from 400 to 850 words
- Use of double passages with two points of view on the same subject
- Introductory and contextual information for the reading passages
- Reading questions that emphasize analytical and evaluative skills



- Passage-based questions testing vocabulary in context
- Discrete questions measuring verbal reasoning and vocabulary in context

One of the major difficulties in talking about the SAT of today in print or in public forums is that critical reading does not fit neatly into a sound bite or a sidebar in a news magazine. To talk about critical reading and what the SAT measures, one has to take the time to do some reading. Here is the text of a recent SAT reading passage:

*The following passage, taken from a book written in 1992, discusses the relative ease with which people can discern meaning from maps.*

The eye and the brain seem to be particularly felicitous partners in the act of map-reading. It is as if we are physiologically disposed to extract information from maps more rapidly, more intuitively, more globally than from, for example, a text or visual scene. That process of visual mining begins with perception—a process that touches on both the physiological and the conceptual processing of map knowledge. Bearing that in mind, we might take a walk with astronomer Patrick Thaddeus, removing him from his preferred milieu, which is mapping carbon monoxide molecules in the Milky Way with a radio telescope at Harvard University, and placing him in a rather less exotic environment—namely, the woods surrounding his country home in upstate New York.

“The forest goes on for miles and miles,” Thaddeus explains. “And I love just walking through the woods by myself. You’re not alone, in the sense that the forest is crisscrossed with deer trails. These deer trails are quite imperceptible. But after a while you know how to recognize them and you can see them. They’re just very faint patterns that generally tend to go in a straight line. Now I followed one of these trails for a mile through the woods. And I suddenly stopped and asked myself, ‘How do I know I’m on this trail?’ But I *am* on it, and I suddenly get shaken off. The signal-to-noise ratio [the relevant information, or ‘signal,’ compared to irrelevant information, or noise] must be one in a thousand, or much less than that. That is, I know I’m on the trail because of a little leaf here, a very faint linear line. But there are much *stronger* sources of noise. Trees across the path, great rocks, and things like that—no computer in the world could possibly filter out that path from all of the conflicting signals around.”

Thaddeus can do this, he believes, because of evolution. “Finding your way home, getting back to your

babies, your families, is something which we and our ancestors, both human and animal, have had to do for not just millions but *tens* of millions of years,” he continues. “Animals are *astonishingly* adept at that, following both visual traces and smell. Smell in humans is a very atrophied sense, but we’re particularly good at visual recognition. So it is technically true that I can follow these trails with a high degree of confidence, where I don’t think any computer in the world has ever been constructed, or could be programmed, to filter out all the noise and not lock onto the tree trunk or things like that. The point is, human beings think in terms of images, and they know what they are looking for. The educated eye knows what it’s looking for, can see things that are, in the technical sense of signal to noise, way, way below one. A very weak, astonishingly weak signal. That is, the human brain is an *incredible* filter for extracting information from confusion.”

Confusion is another name for the world unfiltered, and maps are external, constructed filters that make sense of the confusion, just as the eye and brain are internal, physiological filters that cut through the bewildering mix of signal and noise in a visual scene. By breaking down the graphic or pictorial vocabulary to a bare minimum, maps achieve a visual minimalism that, physiologically speaking, is easy on the eyes. They turn numbers into visual images, create pattern out of measurements, and thus engage the highly evolved human capacity for pattern recognition. Some of the most intense research in the neurosciences today is devoted to elucidating what are described as maps of perception: how perception filters and maps the relentless torrent of information provided by the sense organs, our biotic instruments of measurement. Maps *enable* humans to use inherent biological skills of perception, their “educated” eyes, to separate the message from the static, to see the story line running through random pattern.

This is a complex and challenging text, but a text that is very *interesting*. It has voice, it is original, and it presents ideas that will be unfamiliar to most American high school students in a way that they should be able to understand. It is the kind of thing that can actually change the way you think—if you have never thought about this sort of thing before, it will change the way you look at a forest. If you ask yourself what you want an incoming college freshman to be able to do, presumably the ability to think critically about texts like this would be high on the list.

This passage was followed by 10 critical reading questions. Here is one that refers to the first lines of the passage:

Taking the reader on a “walk” (line 6) primarily serves to

- (A) provide a vicarious experience of moving through space
- (B) make a hypothesis more concrete through a narrative
- (C) demonstrate the ease with which anyone can create a map
- (D) increase respect for the science of astronomical mapping
- (E) suggest the irony of an astronomer’s becoming lost in the woods

(Answer: B)

In this question the test-taker is asked to think about why the writer chose to present his or her argument in a certain way. This is a high-level skill, and the ability to answer this question demonstrates an ability to understand not only what is said but why and how it is said.

The 1994 redesign of the SAT took seriously the idea that changes in the test should have a positive influence on education and that a major task of students in college is to read critically. This modification responded to a 1990 recommendation of the Commission on New Possibilities for the Admissions Testing Program to “approximate more closely the skills used in college and high school work” (*Beyond Prediction*, page 5). At one point consideration was given to redesigning the verbal test to consist entirely of passage-based reading questions. However, the time-consuming nature of these questions made it difficult to design a test with sufficient reliability and other technical characteristics comparable to earlier versions of the SAT.

Table 2 shows how the format of the verbal portion changed between 1958 and 2001.

TABLE 2

**Changes to Format of SAT-Verbal 1958–2001**

	1958–1974	1974–1978	1978–1994	1994–2001
Antonyms	18	25	25	
Analogies	19	20	20	19
Sentence Completions	18	15	15	19
Reading Comprehension	35	25	25	40
Critical Reading	7 passages	5 passages	6 passages	4 passages
Total Verbal	90	85	85	78
Total Testing Time	75 minutes	75 minutes	60 minutes	75 minutes

## IV. Changes to the Mathematical Portion of the SAT Since 1930

The SAT tests given in 1928 and 1929 and between 1936 and 1941 did not contain any mathematics questions. The math section of the SAT administered between 1930 and 1935 contained only free-response questions, and students were given 100 questions to solve in 80 minutes.

The directions from a 1934 math subtest stated: “Write the answer to these questions as quickly as you can. In solving the problems on geometry, use the information given and your own judgment on the geometrical properties of the figures to which you are referred.” Here are two questions from that test:

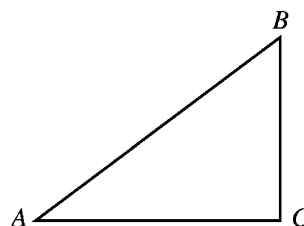


Figure 1

1. In Figure 1, if  $AC = 4$ ,  $BC = 3$ ,  $AB =$   
(Answer:  $AB = 5$ )
2. If  $\frac{b}{2} + \frac{b}{5} = 14$ ,  $b =$   
(Answer:  $b = 20$ )

These questions are straightforward but are not as precise as those written today. In the first question, students were expected to assume that the measure of  $\angle C$  was  $90^\circ$  because the angle looked like a right angle. The only way to find  $AB$  was to use the Pythagorean theorem assuming that  $\triangle ABC$  was a right triangle. The primary challenge of these early tests was mental quickness: How many questions could the student answer correctly in a brief period of time? (Braswell, 1978)

Beginning in 1942 math content on the SAT was tested through the traditional multiple-choice question followed by five choices. The following item is from a 1943 test.

- If  $4b + 2c = 4$ ,  $8b - 2c = 4$ ,  $6b - 3c = (?)$   
 (a) -2 (b) 2 (c) 3 (d) 6 (e) 10

The solution to this problem involves solving simultaneous equations, finding values for  $b$  and  $c$ , and then substituting these values into the expression  $6b - 3c$ .

In 1959 a new math question type (data sufficiency) was introduced. Then in 1974 the data sufficiency questions were replaced with quantitative comparisons, after studies showed that those types of questions had strong predictive validity and were time efficient. Quantitative comparison questions were determined to be somewhat easier to coach students for than others, but they were introduced into the test with the goal of avoiding specific types of questions that were most suitable for pretest coaching.

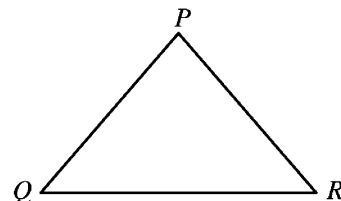
Both the data sufficiency and quantitative comparison questions have answer choices that are the same for all questions. However, the data sufficiency answer choices are much more involved, as the following two examples illustrate. (The directions for the quantitative comparison questions are from an early version. Current directions are shown later.)

### Data Sufficiency Item

*Directions:* Each of the questions below is followed by two statements, labeled (1) and (2), in which certain data are given. In these questions you do not actually have to compute an answer, but rather 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), you are to blacken the space on the answer sheet under

- 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 is sufficient by itself 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.

*Example:*



Can the size of angle  $P$  be determined?

- (1)  $PQ = PR$
- (2) Angle  $Q = 40^\circ$

*Explanation:*

Since  $PQ = PR$  from statement (1),  $\triangle PQR$  is isosceles. Therefore  $\angle Q = \angle R$ .

Since  $\angle Q = 40^\circ$  from statement (2),  $\angle R = 40^\circ$ . It is known that  $\angle P + \angle Q + \angle R = 180^\circ$ . Angle  $P$  can be found by substituting the values of  $\angle Q$  and  $\angle R$  in this equation. Since the problem can be solved and both statements (1) and (2) are needed, the answer is C.

### Quantitative Comparison Item

*Directions:* Each of the following questions 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 greater;
- B if the quantity in Column B is greater;
- C if the two quantities are equal;
- D if the relationship cannot be determined from the information given.

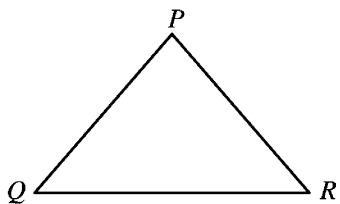
*Notes:*

1. In certain questions, information concerning one or both of the quantities to be compared is centered above the two columns.
2. A symbol that appears in both columns represents the same thing in Column A as it does in Column B.
3. Letters such as  $x$ ,  $n$ , and  $k$  stand for real numbers.

EXAMPLES			
	<i>Column A</i>	<i>Column B</i>	<i>Answers</i>
E 1.	$2 \times 6$	$2 + 6$	<input checked="" type="radio"/> (A) <input type="radio"/> (B) <input type="radio"/> (C) <input type="radio"/> (D)
E 2.	$180 - x$	$y$	<input type="radio"/> (A) <input type="radio"/> (B) <input checked="" type="radio"/> (C) <input type="radio"/> (D)
E 3.	$p - q$	$q - p$	<input type="radio"/> (A) <input type="radio"/> (B) <input checked="" type="radio"/> (C) <input type="radio"/> (D)

Example:  
Column A

Column B



Note: Figure not drawn to scale.

$$PQ = PR$$

The measure of  $\angle Q$

The measure of  $\angle P$

*Explanation:*

Since  $PQ = PR$ , the measure of  $\angle Q$  equals the measure of  $\angle R$ . They could both equal  $40^\circ$ , in which case the measure of  $\angle P$  would equal  $100^\circ$ . The measure of  $\angle Q$  and the measure of  $\angle R$  could both equal  $80^\circ$ , in which case the measure of  $\angle P$  would equal  $20^\circ$ . In one case, the measure of  $\angle Q$  would be less than the measure of  $\angle P$  ( $40^\circ < 100^\circ$ ). In the other case, the measure of  $\angle Q$  would be greater than the measure of  $\angle P$  ( $80^\circ > 20^\circ$ ). Therefore, the answer to this question is (D) since a relationship cannot be determined from the information given.

TABLE 3

Changes to Format of SAT-Math 1942–2001

	1942–1959	1959–1974	1974–1994	1994–2001
5-Choice Multiple Choice	48	42	40	35
Data Sufficiency	12	18		
Quantitative Comparison			20	15
Student-Produced Response				10
Total Mathematical Items	60	60	60	60
Total Testing Time	75 minutes	75 minutes	60 minutes	75 minutes

Note that both questions test similar math content, but the quantitative comparison question takes much less time to solve and is less dependent on verbal skills than is the data sufficiency question. Quantitative comparison questions have been found to be generally more appropriate for disadvantaged students than data sufficiency items (Braswell, 1978).

Two major changes to the math section of the SAT took place in 1994: the inclusion of some questions that require test-takers to produce their own solutions rather than select multiple-choice alternatives, and a policy permitting the use of calculators.

Table 3 shows how the format of the math portion of the test has changed between 1942 and 2001.

The 1994 changes were made for a variety of reasons (Braswell, 1991); three very important ones were to:

- Strengthen the relationship between the test and current mathematics curriculum
- Move away from an exclusively multiple-choice test
- Reduce the impact of speed on test performance

An important impetus for change was that the National Council of Teachers of Mathematics (NCTM) had suggested increased attention in the mathematics curriculum to the use of real-world problems; probability and statistics; problem solving, reasoning, and analyzing; application of learning to new contexts; and solving problems that were not multiple-choice (including problems that had more than one answer). This group also strongly encouraged permitting the use of calculators on the test.

The 1994 changes were responsive to NCTM suggestions. Since then there has been a concerted effort to avoid contrived word problems and to include real-world problems that may be more interesting and have meaning to students. Here are two real-world problems from more recent tests:

An aerobics instructor burns 3,000 calories per day for 4 days. How many calories must she burn during the next day so that the average (arithmetic mean) number of calories burned for the 5 days is 3,500 calories per day?

- (A) 6,000
- (B) 5,500
- (C) 5,000
- (D) 4,500
- (E) 4,000

(Answer: B)

A certain building has 2,600 square feet of surface that needs to be painted. If 1 gallon of paint will cover 250 square feet, what is the least whole number of gallons that must be purchased in order to have enough paint to apply one coat to the surface? (Assume that only whole gallons of paint can be purchased.)

- (A) 5
- (B) 10
- (C) 11
- (D) 15
- (E) 110

(Answer: C)

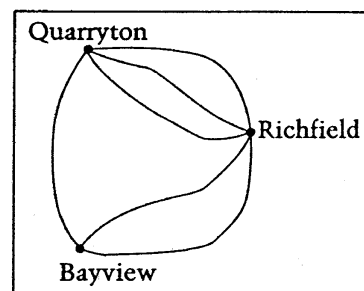
The specifications changed in 1994 to require probability, elementary statistics, and counting problems on each test. Concepts of median and mode were also introduced.

20, 30, 50, 70, 80, 80, 90

Seven students played a game and their scores from least to greatest are give above. Which of the following is true of the scores?

- I. The average (arithmetic mean) is greater than 70.
  - II. The median is greater than 70.
  - III. The mode is greater than 70.
- (A) None
  - (B) III only
  - (C) I and II only
  - (D) II and III only
  - (E) I, II, and III

(Answer: B)



The figure above shows all roads between Quarryton, Richfield, and Bayview. Martina is traveling from Quarryton to Bayview *and back*. How many different ways could she make the round-trip, going through Richfield exactly once on a round-trip and not traveling any section of road more than once on a round-trip?

- (A) 5
- (B) 6
- (C) 10
- (D) 12
- (E) 16

(Answer: D)

Math sections of the SAT focus on quantitative reasoning and problem solving. The test does not measure skills in advanced math, but it does challenge students to apply strong problem-solving techniques and use the math they do know in flexible and creative ways. The test demands that students go beyond applying rules and formulas to think through problems they haven't solved before.

### *Student-Produced Response Questions*

Student-produced response (SPR) questions were also added to the test in 1994 in response to the NCTM Standards.

The SPR format has many advantages:

- It eliminates guessing and back-door approaches that depend on answer choices.
- The statistical data are more reliable (there is almost no guessing).
- The grid used to record the answer accommodates different forms of the correct answer (fraction versus decimal).
- It allows questions that have more than one correct answer.

### Directions for Student-Produced Response Questions

Each of the remaining 8 questions (33-40) 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  $7/12$

Write answer in boxes. →

7	/	1	2
○	○	○	○
○	○	○	○
○	○	○	○
○	○	○	○
○	○	○	○
○	○	○	○
○	○	○	○
○	○	○	○
○	○	○	○
○	○	○	○

← Fraction line

Grid in result. →

Answer: 2.5

2	.	5	
○	○	○	○
○	○	○	○
○	○	○	○
○	○	○	○
○	○	○	○
○	○	○	○
○	○	○	○
○	○	○	○
○	○	○	○
○	○	○	○

← Decimal point

Answer: 201  
Either position is correct.

2	0	1	
○	○	○	○
○	○	○	○
○	○	○	○
○	○	○	○
○	○	○	○
○	○	○	○
○	○	○	○
○	○	○	○
○	○	○	○
○	○	○	○

2	0	1	
○	○	○	○
○	○	○	○
○	○	○	○
○	○	○	○
○	○	○	○
○	○	○	○
○	○	○	○
○	○	○	○
○	○	○	○
○	○	○	○

**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  $\frac{21}{12}$  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  $\frac{2}{3} = .6666 \dots$

2	/	3	
○	○	○	○
○	○	○	○
○	○	○	○
○	○	○	○
○	○	○	○
○	○	○	○
○	○	○	○
○	○	○	○
○	○	○	○
○	○	○	○

.	6	6	6
○	○	○	○
○	○	○	○
○	○	○	○
○	○	○	○
○	○	○	○
○	○	○	○
○	○	○	○
○	○	○	○
○	○	○	○
○	○	○	○

.	6	6	7
○	○	○	○
○	○	○	○
○	○	○	○
○	○	○	○
○	○	○	○
○	○	○	○
○	○	○	○
○	○	○	○
○	○	○	○
○	○	○	○

Student-produced response questions test reasoning skills that could not be tested as effectively in a multiple-choice format, as illustrated by the following example.

What is the greatest 3-digit integer that is a multiple of 10?  
(Answer: 990)

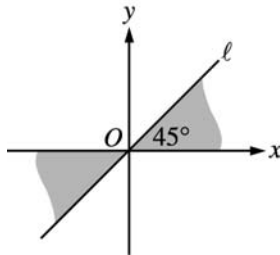
There is reasoning involved in determining that 990 is the answer to this question. This would be a trivial problem if answer choices were given.

The SPR format also allows for questions with more than one answer. The following problem is an example of a question with a set of discrete answers.

The sum of  $k$  and  $k + 1$  is greater than 9 but less than 17. If  $k$  is an integer, what is one possible value of  $k$ ?

Solving the inequality  $9 < k + (k + 1) < 17$  yields  $4 < k < 8$ . Since  $k$  is an integer, the answer to this question could be 5, 6, or 7. Students may grid any of these three integers as an answer.

Another type of SPR question has correct answers in a range. The answer to the following question involving the slope of a line is any number between 0 and 1. Students may grid any number in the interval between 0 and 1 that the grid can accommodate— $\frac{1}{2}$ , .001, .98, etc. Slope was another topic added to the SAT in 1994 because of its increased importance in the curriculum.



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

The introduction of calculator use on the math portion of the test reflected changes in the use of calculators in mathematics instruction. The following quantitative comparison question was used in the SAT before calculator use was permitted, but it is no longer appropriate for the test. The directions that precede the examples

are in the form currently used on the test (See directions on the bottom of this page.)

Column A

$$3 \times 352 \times 8$$

Column B

$$4 \times 352 \times 6$$

Explanation:

Since 352 appears in the product in both Column A and Column B, it is only necessary to compare  $3 \times 8$  with  $4 \times 6$ . These products are equal, so the answer to this problem is (C).

This question tested reasoning when calculator use was not permitted, but it only tests button pushing when calculators are allowed. A more appropriate question for a current SAT would be:

Column A

$$0 < x < 1 < y < z$$

$$2xy$$

Column B

$$2yz$$

This question invites a comparison of two products, and since both products contain  $2y$ , and  $y > 0$ , it is only necessary to compare  $x$  with  $z$ . Since  $x < z$ , the correct answer is (B), as the quantity in column B is greater than the quantity in column A.

### Directions for Quantitative Comparison Questions

Questions 21-32 each consist of two quantities in boxes, one in Column A and one in Column B. You are to compare the two quantities and on the answer sheet fill in oval

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

#### Notes:

1. In some questions, information is given about one or both of the quantities to be compared. In such cases, the given information is centered above the two columns and is not boxed.
2. In a given question, a symbol that appears in both columns represents the same thing in Column A as it does in Column B.
3. Letters such as  $x$ ,  $n$ , and  $k$  stand for real numbers.

#### EXAMPLES

Column A

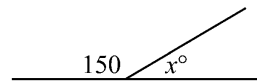
E1  $5^2$

Column B

20

Answers

- (A)  (B)  (C)  (D)



E2  $x$

30

- (A)  (B)  (C)  (D)

$r$  and  $s$  are integers.

E3  $r + 1$

$s - 1$

- (A)  (B)  (C)  (D)

---

## V. What the SAT Measures Today

This historical perspective has attempted to show how the SAT has evolved over time. As can be seen from the narrative above, the SAT given today is quite different from the test given as recently as 10 years ago.

The verbal portion of today's SAT can be described as a measure of the fundamental academic skill of constructing meaning out of the English language in such a way as to be able to understand and participate in certain kinds of formal discourse. This section of the test focuses primarily on critical reading. Students are asked to read passages from the sciences, the social sciences, and the humanities, and to reflect on the author's point of view, technique, and logic. Critical reading skills are increasingly important for success in high school. Most high school exit exams in English focus on critical reading of challenging nonfiction.

The math portion of today's SAT can be described as a measure of the ability to use mathematical concepts and skills in order to engage in problem solving. The test does not measure advanced math skills such as trigonometry or calculus. But it does challenge students to apply strong problem-solving techniques and use the math they do know in flexible ways. It asks that students go beyond applying rules and formulas to think through problems they have not solved before. This emphasis on problem solving in mathematics mirrors the higher academic standards that are in effect in virtually every state. The National Council of Teachers of Mathematics and other bodies have long argued that mathematics education should not merely inculcate students with knowledge of facts and algorithms but should aim to create flexible thinkers who are comfortable handling nonroutine problems.

Clearly, the SAT is a demanding test that focuses on assessing fundamental math and reading skills that are crucial to success in college and adult life. Although the SAT is not designed to be a measure of the quality of one's high school education, the reading and problem-solving skills measured by the test are certainly consistent with school-based learning, and these academic skills are very much related to the skills needed to succeed in college.

## VI. Conclusion

This paper has shown the various ways in which the SAT has evolved since its introduction in 1926. Some of the modifications have involved changes in the types of ques-

tions used to measure verbal and mathematical skills. Other modifications focused on liberalizing time limits to ensure that speed of responding to questions has minimal effect on performance. There were changes in the actual administration of the test, such as allowing students to use calculators on the math sections. Still other revisions have stemmed from a concern that certain types of questions might be more susceptible to pretest coaching. All of these changes were intended to update the SAT so that it remains fair for an increasingly diverse group of test-takers, while at the same time enhancing its effectiveness as an admissions tool. The most recent changes were also heavily influenced by a desire to reflect contemporary secondary school curriculum and reinforce sound educational standards and practices.

Reviewing the content and format of the test over its 75-year history, it is logical to ask whether and how the test might change in the future. Given the evolutionary past of the SAT, it seems highly probable that it will continue to be modified. The SAT Committee, an external advisory panel that includes high school and university faculty and administrators, routinely considers possible changes. Among the topics that have been discussed in recent years by the SAT Committee are exploring ways to further liberalize the amount of time students have to complete the test; possible elimination or replacement of the current analogy question type; changing from a formula-based to a rights-only scoring system; possible prohibition of calculators on certain questions or prohibition of certain types of calculators; and adopting guidelines on the types of vocabulary that are permissible.

Many of the motivations that led to previous modifications in the SAT continue to be relevant. The test must continue to provide useful information for admissions purposes. It must continue to be fair to all students who take it. It must be as impervious to special coaching efforts as possible (although it must be recognized that there is a point at which good coaching and good education intersect). Future changes will also need to reflect the latest trends in secondary education reform and should have a positive (or at least neutral) effect on students' academic growth.

A modified test should have all of the strong technical qualities of the current test (i.e., appropriately timed with a high level of measurement precision across the full range of test-takers and equivalent predictive validity). The meaning of the score scale should not change. A revised test should be designed so that its scores can be linked to scores on the present test.

Although there could be many reasons for changing the content and/or format of the test, the basic and most important challenge is always to ensure that the SAT is the fairest test possible for students and that it effectively meets the needs of college admissions offices.



---

## References

- Braswell, James, "The College Board Scholastic Aptitude Test: An Overview of the Mathematical Portion," *The Mathematics Teacher*, March 1978.
- Braswell, James, "Overview of Changes in the SAT Mathematics Test in 1994," paper presented at the annual meeting of the National Council on Measurement in Education, April 5, 1991, Chicago.
- Coffman, William E., "The Scholastic Aptitude Test 1926–1962," paper presented to the Committee of Examiners on Aptitude Testing, 1962.
- Commission on New Possibilities for the Admissions Testing Program, *Beyond Prediction* (New York: College Entrance Examination Board, 1990).
- Curley, W. Edward, and Geraldine May, "Content Rationale for the New SAT-Verbal," paper presented at the annual meeting of the National Council on Measurement in Education, April 5, 1991, Chicago.
- Loret, Peter G., *A History of the Content of the Scholastic Aptitude Test* (Princeton, N.J.: Educational Testing Service, 1960).





