Test

What You Need to Know
SAT

The SAT is a standardized test for college admissions in the United States. The SAT is owned, published, and developed by the College Board, a nonprofit organization in the United States. It was formerly developed, published, and scored by the Educational Testing Service[1] which still administers the exam. The test is intended to assess a student's readiness for college. It was first introduced in 1926, and its name and scoring have changed several times. It was first called the Scholastic Aptitude Test, then the Scholastic Assessment Test, but now SAT does not stand for anything, hence it is an empty acronym.

The current SAT Reasoning Test, introduced in 2005, takes three hours and forty-five minutes to finish, and costs $50 ($81 International), excluding late fees.[2] Possible scores range from 600 to 2400, combining test results from three 800-point sections (Mathematics, Critical Reading, and Writing).

Taking the SAT or its competitor, the ACT, is required for freshman entry to many, but not all, universities in the United States.[3]

Function

The College Board states that SAT measures literacy and writing skills that are needed for academic success in college. They state that the SAT assesses how well the test takers analyze and solve problems—skills they learned in school that they will need in college. The SAT is typically taken by high school sophomores, juniors and seniors.[4] Specifically, the College Board states that use of the SAT in combination with high school grade point average (GPA) provides a better indicator of success in college than high school grades alone, as measured by college freshman GPA. Various studies conducted over the lifetime of the SAT show a statistically significant increase in correlation of high school grades and freshman grades when the SAT is factored in.[5]

There are substantial differences in funding, curricula, grading, and difficulty among U.S. secondary schools due to American federalism, local control, and the prevalence of private, distance, and home schooled students. SAT (and ACT) scores are intended to supplement the secondary school record and help admission officers put local data—such as course work, grades, and class rank—in a national perspective.[6]

Historically, the SAT has been more popular among colleges on the coasts and the ACT more popular in the Midwest and South. There are some colleges that require the ACT to be taken for college course placement, and a few schools that formerly did not accept the SAT at all. Nearly all colleges accept the test.[7]

Certain high IQ societies, like Mensa, the Prometheus Society and the Triple Nine Society, use scores from certain years as one of their admission tests. For instance, the Triple Nine Society accepts scores of 1450 on tests taken before April 1995, and scores of at least 1520 on tests taken between April 1995 and February 2005.[8]

The SAT is sometimes given to students younger than 13 by organizations such as the Study of Mathematically Precocious Youth, who use the results to select, study and mentor students of exceptional ability.

While the exact manner in which SAT scores will help to determine admission of a student at American institutions of higher learning is generally a matter decided by the individual institution, some foreign countries have made SAT (and ACT) scores a legal criterion in deciding whether holders of American high school diplomas will be admitted at their public universities.
Structure

SAT consists of three major sections: Critical Reading, Mathematics, and Writing. Each section receives a score on the scale of 200–800. All scores are multiples of 10. Total scores are calculated by adding up scores of the three sections. Each major section is divided into three parts. There are 10 sub-sections, including an additional 25-minute experimental or "equating" section that may be in any of the three major sections. The experimental section is used to normalize questions for future administrations of the SAT and does not count toward the final score. The test contains 3 hours and 45 minutes of actual timed sections, most administrations (after including orientation, distribution of materials, completion of biographical sections, and eleven minutes of timed breaks) run for about four and a half hours. The questions range from easy, medium, and hard depending on the scoring from the experimental sections. Easier questions typically appear closer to the beginning of the section while harder questions are towards the end in certain sections. This is not true for every section (the Critical Reading section is in chronological order) but it is the rule of thumb mainly for math and the 19 sentence completions on the test.

Critical Reading

The Critical Reading (formerly Verbal) section of the SAT is made up of three scored sections: two 25-minute sections and one 20-minute section, with varying types of questions, including sentence completions and questions about short and long reading passages. Critical Reading sections normally begin with 5 to 8 sentence completion questions; the remainder of the questions are focused on the reading passages. Sentence completions generally test the student’s vocabulary and understanding of sentence structure and organization by requiring the student to select one or two words that best complete a given sentence. The bulk of the Critical Reading section is made up of questions regarding reading passages, in which students read short excerpts on social sciences, humanities, physical sciences, or personal narratives and answer questions based on the passage. Certain sections contain passages asking the student to compare two related passages; generally, these consist of shorter reading passages. The number of questions about each passage is proportional to the length of the passage. Unlike in the Mathematics section, where questions go in the order of difficulty, questions in the Critical Reading section go in the order of the passage. Overall, question sets towards the beginning of the section are easier, and question sets towards the end of the section are harder.

Mathematics

The Mathematics section of the SAT is widely known as the Quantitative Section or Calculation Section. The mathematics section consists of three scored sections. There are two 25-minute sections and one 20-minute section, as follows:

- One of the 25-minute sections is entirely multiple choice, with 20 questions.
- The other 25-minute section contains 8 multiple choice questions and 10 grid-in questions. For grid-in questions, test-takers write the answer inside a grid on the answer sheet. Unlike multiple choice questions, there is no penalty for incorrect answers on grid-in questions because the test-taker is not limited to a few possible choices.
- The 20-minute section is all multiple choice, with 16 questions.

The SAT has done away with quantitative comparison questions on the math section, leaving only questions with symbolic or numerical answers.
• New topics include Algebra II and scatter plots. These recent changes have resulted in a shorter, more quantitative exam requiring higher level mathematics courses relative to the previous exam.

**Calculator use**

Four-function, scientific, and graphing calculators are permitted on the SAT math section; however, calculators are not permitted on either of the other sections. Calculators with QWERTY keyboards, cell phone calculators, portable computers, and personal organizers are not permitted.

With the recent changes to the content of the SAT math section, the need to save time while maintaining accuracy of calculations has led some to use calculator programs during the test. These programs allow students to complete problems faster than would normally be possible when making calculations manually.

The use of a graphing calculator is sometimes preferred, especially for geometry problems and questions involving multiple calculations. According to research conducted by the CollegeBoard, performance on the math sections of the exam is associated with the extent of calculator use, with those using calculators on about a third to a half of the items averaging higher scores than those using calculators less frequently. The use of a graphing calculator in mathematics courses, and also becoming familiar with the calculator outside of the classroom, is known to have a positive effect on the performance of students using a graphing calculator during the exam.

**Writing**

The writing section of the SAT, based on but not directly comparable to the old SAT II subject test in writing (which in turn was developed from the old TSWE), includes multiple choice questions and a brief essay. The essay subscore contributes about 28% towards the total writing score, with the multiple choice questions contributing 70%. This section was implemented in March 2005 following complaints from colleges about the lack of uniform examples of a student's writing ability and critical thinking.

The multiple choice questions include error identification questions, sentence improvement questions, and paragraph improvement questions. Error identification and sentence improvement questions test the student's knowledge of grammar, presenting an awkward or grammatically incorrect sentence; in the error identification section, the student must locate the word producing the source of the error or indicate that the sentence has no error, while the sentence improvement section requires the student to select an acceptable fix to the awkward sentence. The paragraph
improvement questions test the student's understanding of logical organization of ideas, presenting a poorly written student essay and asking a series of questions as to what changes might be made to best improve it.

The essay section, which is always administered as the first section of the test, is 25 minutes long. All essays must be in response to a given prompt. The prompts are broad and often philosophical and are designed to be accessible to students regardless of their educational and social backgrounds. For instance, test takers may be asked to expand on such ideas as their opinion on the value of work in human life or whether technological change also carries negative consequences to those who benefit from it. No particular essay structure is required, and the College Board accepts examples "taken from [the student's] reading, studies, experience, or observations." Two trained readers assign each essay a score between 1 and 6, where a score of 0 is reserved for essays that are blank, off-topic, non-English, not written with a Number 2 pencil, or considered illegible after several attempts at reading. The scores are summed to produce a final score from 2 to 12 (or 0). If the two readers' scores differ by more than one point, then a senior third reader decides. The average time each reader/grader spends on each essay is less than 3 minutes.\textsuperscript{[13]}

In March 2004, Dr. Les Perelman analyzed 15 scored sample essays contained in the College Board's ScoreWrite book along with 30 other training samples and found that in over 90% of cases, the essay's score could be predicted from simply counting the number of words in the essay.\textsuperscript{[13]} Two years later, Dr. Perelman trained high school seniors to write essays that made little sense but contained infrequently used words such as "plethora" and "myriad." All of the students received scores of "10" or better, which placed the essays in the 92nd percentile or higher.\textsuperscript{[14]}

### Style of questions

Most of the questions on the SAT, except for the essay and the grid-in math responses, are multiple choice; all multiple-choice questions have five answer choices, one of which is correct. The questions of each section of the same type are generally ordered by difficulty. However, an important exception exists: Questions that follow the long and short reading passages are organized chronologically, rather than by difficulty. Ten of the questions in one of the math sub-sections are not multiple choice. They instead require the test taker to bubble in a number in a four-column grid.

The questions are weighted equally. For each correct answer, one raw point is added. For each incorrect answer one-fourth of a point is deducted.\textsuperscript{[15]} No points are deducted for incorrect math grid-in questions. This ensures that a student's mathematically expected gain from guessing is zero. The final score is derived from the raw score; the precise conversion chart varies between test administrations.

The SAT therefore recommends only making educated guesses, that is, when the test taker can eliminate at least one answer he or she thinks is wrong. Without eliminating any answers one's probability of answering correctly is 20%. Eliminating one wrong answer increases this probability to 25% (and the expected gain to 1/16 of a point); two, a 33.3% probability (1/6 of a point); and three, a 50% probability (3/8 of a point).

<table>
<thead>
<tr>
<th>Section</th>
<th>Average Score</th>
<th>Time (Minutes)</th>
<th>Content</th>
</tr>
</thead>
<tbody>
<tr>
<td>Writing</td>
<td>493</td>
<td>60</td>
<td>Grammar, usage, and diction.</td>
</tr>
<tr>
<td>Mathematics</td>
<td>515</td>
<td>70</td>
<td>Number and operations; algebra and functions; geometry; statistics, probability, and data analysis</td>
</tr>
<tr>
<td>Critical Reading</td>
<td>501</td>
<td>70</td>
<td>Vocabulary, Critical reading, and sentence-level reading</td>
</tr>
</tbody>
</table>
Taking the test

The SAT is offered seven times a year in the United States; in October, November, December, January, March (or April, alternating), May, and June. The test is typically offered on the first Saturday of the month for the November, December, May, and June administrations. In other countries, the SAT is offered on the same dates as in the United States except for the first spring test date (i.e., March or April), which is not offered. In 2006, the test was taken
1,465,744 times.[16]

Candidates may take either the SAT Reasoning Test or up to three SAT Subject Tests on any given test date, except the first spring test date, when only the SAT Reasoning Test is offered. Candidates wishing to take the test may register online at the College Board's website, by mail, or by telephone, at least three weeks before the test date.

The SAT Subject Tests are all given in one large book on test day. Therefore, it is actually immaterial which tests, and how many, the student signs up for; with the possible exception of the language tests with listening, the student may change his or her mind and take any tests, regardless of his or her initial sign-ups. Students who choose to take more subject tests than they signed up for will later be billed by College Board for the additional tests and their scores will be withheld until the bill is paid. Students who choose to take fewer subject tests than they signed up for are not eligible for a refund.

The SAT Reasoning Test costs $49 ($78 International, $99 for India and Pakistan). For the Subject tests, students pay a $22 ($49 International, $73 for India and Pakistan) Basic Registration Fee and $11 per test (except for language tests with listening, which cost $21 each).[2] The College Board makes fee waivers available for low income students. Additional fees apply for late registration, standby testing, registration changes, scores by telephone, and extra score reports (beyond the four provided for free).

Candidates whose religious beliefs prevent them from taking the test on a Saturday may request to take the test on the following day, except for the October test date in which the Sunday test date is eight days after the main test offering. Such requests must be made at the time of registration and are subject to denial.

Students with verifiable disabilities, including physical and learning disabilities, are eligible to take the SAT with accommodations. The standard time increase for students requiring additional time due to learning disabilities is time + 50%; time + 100% is also offered.

Raw scores, scaled scores, and percentiles

Students receive their online score reports approximately three weeks after test administration (six weeks for mailed, paper scores), with each section graded on a scale of 200–800 and two sub scores for the writing section: the essay score and the multiple choice sub score. In addition to their score, students receive their percentile (the percentage of other test takers with lower scores). The raw score, or the number of points gained from correct answers and lost from incorrect answers (ranges from just under 50 to just under 60, depending upon the test), is also included.[17] Students may also receive, for an additional fee, the Question and Answer Service, which provides the student’s answer, the correct answer to each question, and online resources explaining each question.

The corresponding percentile of each scaled score varies from test to test—for example, in 2003, a scaled score of 800 in both sections of the SAT Reasoning Test corresponded to a percentile of 99.9, while a scaled score of 800 in the SAT Physics Test corresponded to the 94th percentile. The differences in what scores mean with regard to percentiles are due to the content of the exam and the caliber of students choosing to take each exam. Subject Tests are subject to intensive study (often in the form of an AP, which is relatively more difficult), and only those who know they will perform well tend to take these tests, creating a skewed distribution of scores.

The percentiles that various SAT scores for college-bound seniors correspond to are summarized in the following chart.[18][19]
<table>
<thead>
<tr>
<th>Percentile</th>
<th>Score, 1600 Scale (official, 2006)</th>
<th>Score, 2400 Scale (official, 2006)</th>
</tr>
</thead>
<tbody>
<tr>
<td>99.93/99.98*</td>
<td>1600</td>
<td>2400</td>
</tr>
<tr>
<td>99+ **</td>
<td>≥1540</td>
<td>≥2280</td>
</tr>
<tr>
<td>99</td>
<td>≥1480</td>
<td>≥2200</td>
</tr>
<tr>
<td>98</td>
<td>≥1450</td>
<td>≥2140</td>
</tr>
<tr>
<td>97</td>
<td>≥1420</td>
<td>≥2100</td>
</tr>
<tr>
<td>93</td>
<td>≥1340</td>
<td>≥1990</td>
</tr>
<tr>
<td>88</td>
<td>≥1280</td>
<td>≥1900</td>
</tr>
<tr>
<td>81</td>
<td>≥1220</td>
<td>≥1800</td>
</tr>
<tr>
<td>72</td>
<td>≥1150</td>
<td>≥1700</td>
</tr>
<tr>
<td>61</td>
<td>≥1090</td>
<td>≥1600</td>
</tr>
<tr>
<td>48</td>
<td>≥1010</td>
<td>≥1500</td>
</tr>
<tr>
<td>36</td>
<td>≥950</td>
<td>≥1400</td>
</tr>
<tr>
<td>24</td>
<td>≥870</td>
<td>≥1300</td>
</tr>
<tr>
<td>15</td>
<td>≥810</td>
<td>≥1200</td>
</tr>
<tr>
<td>8</td>
<td>≥730</td>
<td>≥1090</td>
</tr>
<tr>
<td>4</td>
<td>≥650</td>
<td>≥990</td>
</tr>
<tr>
<td>2</td>
<td>≥590</td>
<td>≥890</td>
</tr>
</tbody>
</table>

* The percentile of the perfect score was 99.98 on the 2400 scale and 99.93 on the 1600 scale.

** 99+ means better than 99.5 percent of test takers.

The older SAT (before 1995) had a very high ceiling. In any given year, only seven of the million test-takers scored above 1580. A score above 1580 was equivalent to the 99.9995 percentile. [20]

**SAT-ACT score comparisons**

Although there is no official conversion chart between the SAT and its biggest rival, the ACT, the College Board released an unofficial chart based on results from 103,525 test takers who took both tests between October 1994 and December 1996;[21] however, both tests have changed since then. Several colleges have also issued their own charts. The following is based on the University of California's conversion chart.[22]
### Correlations with IQ

Frey and Detterman (2003) analyzed the correlation of SAT scores with intelligence test scores.\(^{[23]}\) They found SAT scores to be highly correlated with general mental ability, or \(g\) (\(r=\cdot82\) in their sample, .86 when corrected for non-linearity). The correlation between SAT scores and scores on the Raven's Advanced Progressive Matrices was .483 (.72 corrected for restricted range). They concluded that the SAT is primarily a test of \(g\). Beaujean and colleagues (2006) have reached similar conclusions.\(^{[24]}\)
## History

**Mean SAT Scores by year**[^25]

<table>
<thead>
<tr>
<th>Year of exam</th>
<th>Reading/Verbal Score</th>
<th>Math Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1972</td>
<td>530</td>
<td>509</td>
</tr>
<tr>
<td>1973</td>
<td>523</td>
<td>506</td>
</tr>
<tr>
<td>1974</td>
<td>521</td>
<td>505</td>
</tr>
<tr>
<td>1975</td>
<td>512</td>
<td>498</td>
</tr>
<tr>
<td>1976</td>
<td>509</td>
<td>497</td>
</tr>
<tr>
<td>1977</td>
<td>507</td>
<td>496</td>
</tr>
<tr>
<td>1978</td>
<td>507</td>
<td>494</td>
</tr>
<tr>
<td>1979</td>
<td>505</td>
<td>493</td>
</tr>
<tr>
<td>1980</td>
<td>502</td>
<td>492</td>
</tr>
<tr>
<td>1981</td>
<td>502</td>
<td>492</td>
</tr>
<tr>
<td>1982</td>
<td>504</td>
<td>493</td>
</tr>
<tr>
<td>1983</td>
<td>503</td>
<td>494</td>
</tr>
<tr>
<td>1984</td>
<td>504</td>
<td>497</td>
</tr>
<tr>
<td>1985</td>
<td>509</td>
<td>500</td>
</tr>
<tr>
<td>1986</td>
<td>509</td>
<td>500</td>
</tr>
<tr>
<td>1987</td>
<td>507</td>
<td>501</td>
</tr>
<tr>
<td>1988</td>
<td>505</td>
<td>501</td>
</tr>
<tr>
<td>1989</td>
<td>504</td>
<td>502</td>
</tr>
<tr>
<td>1990</td>
<td>500</td>
<td>501</td>
</tr>
<tr>
<td>1991</td>
<td>499</td>
<td>500</td>
</tr>
<tr>
<td>1992</td>
<td>500</td>
<td>501</td>
</tr>
<tr>
<td>1993</td>
<td>500</td>
<td>503</td>
</tr>
<tr>
<td>1994</td>
<td>499</td>
<td>504</td>
</tr>
<tr>
<td>1995</td>
<td>504</td>
<td>506</td>
</tr>
<tr>
<td>1996</td>
<td>505</td>
<td>508</td>
</tr>
<tr>
<td>1997</td>
<td>505</td>
<td>511</td>
</tr>
<tr>
<td>1998</td>
<td>505</td>
<td>512</td>
</tr>
<tr>
<td>1999</td>
<td>505</td>
<td>511</td>
</tr>
<tr>
<td>2000</td>
<td>505</td>
<td>514</td>
</tr>
<tr>
<td>2001</td>
<td>506</td>
<td>514</td>
</tr>
<tr>
<td>2002</td>
<td>504</td>
<td>516</td>
</tr>
<tr>
<td>2003</td>
<td>507</td>
<td>519</td>
</tr>
<tr>
<td>2004</td>
<td>508</td>
<td>518</td>
</tr>
<tr>
<td>2005</td>
<td>508</td>
<td>520</td>
</tr>
</tbody>
</table>

[^25]: Information from a specific source is not provided.
Originally used mainly by colleges and universities in the northeastern United States, and developed by Carl Brigham, one of the psychologists who worked on the Army Alpha and Beta tests, the SAT was originally developed as a way to eliminate test bias between people from different socio-economic backgrounds.

### 1901 test

The College Board began on June 17, 1901, when 973 students took its first test, across 67 locations in the United States, and two in Europe. Although those taking the test came from a variety of backgrounds, approximately one third were from New York, New Jersey, or Pennsylvania. The majority of those taking the test were from private schools, academies, or endowed schools. About 60% of those taking the test applied to Columbia University. The test contained sections on English, French, German, Latin, Greek, history, mathematics, chemistry, and physics. The test was not multiple choice, but instead was evaluated based on essay responses as "excellent", "good", "doubtful", "poor" or "very poor".\[26\]

### 1926 test

The first administration of the SAT occurred on June 23, 1926, when it was known as the Scholastic Aptitude Test.\[27\][28] This test, prepared by a committee headed by Princeton psychologist Carl Campbell Brigham, had sections of definitions, arithmetic, classification, artificial language, antonyms, number series, analogies, logical inference, and paragraph reading. It was administered to over 8,000 students at over 300 test centers. Men composed 60% of the test-takers. Slightly over a quarter of males and females applied to Yale University and Smith College.\[28\] The test was paced rather quickly, test-takers being given only a little over 90 minutes to answer 315 questions.\[27\]

### 1928 and 1929 tests

In 1928 the number of verbal sections was reduced to 7, and the time limit was increased to slightly under two hours. In 1929 the number of sections was again reduced, this time to 6. These changes in part loosened time constraints on test-takers. Math was eliminated entirely for these tests, instead focusing only on verbal ability.\[27\]

### 1930 test and 1936 changes

In 1930 the SAT was first split into the verbal and math sections, a structure that would continue through 2004. The verbal section of the 1930 test covered a more narrow range of content than its predecessors, examining only antonyms, double definitions (somewhat similar to sentence completions), and paragraph reading. In 1936, analogies were re-added. Between 1936 and 1946, students had between 80 and 115 minutes to answer 250 verbal questions (over a third of which were on antonyms). The mathematics test introduced in 1930 contained 100 free response questions to be answered in 80 minutes, and focused primarily on speed. From 1936 to 1941, like the 1928 and 1929 tests, the mathematics section was eliminated entirely. When the mathematics portion of the test was re-added in

### Mean SAT Scores by year

<table>
<thead>
<tr>
<th>Year</th>
<th>Reading</th>
<th>Math</th>
</tr>
</thead>
<tbody>
<tr>
<td>2006</td>
<td>503</td>
<td>518</td>
</tr>
<tr>
<td>2007</td>
<td>502</td>
<td>515</td>
</tr>
<tr>
<td>2008</td>
<td>502</td>
<td>515</td>
</tr>
<tr>
<td>2009</td>
<td>501</td>
<td>515</td>
</tr>
<tr>
<td>2010</td>
<td>501</td>
<td>516</td>
</tr>
<tr>
<td>2011</td>
<td>497</td>
<td>514</td>
</tr>
</tbody>
</table>
1942, it consisted of multiple choice questions.[27]

1946 test and associated changes
Paragraph reading was eliminated from the verbal portion of the SAT in 1946, and replaced with reading comprehension, and "double definition" questions were replaced with sentence completions. Between 1946 and 1957 students were given 90 to 100 minutes to complete 107 to 170 verbal questions. Starting in 1958 time limits became more stable, and for 17 years, until 1975, students had 75 minutes to answer 90 questions. In 1959 questions on data sufficiency were introduced to the mathematics section, and then replaced with quantitative comparisons in 1974. In 1974 both verbal and math sections were reduced from 75 minutes to 60 minutes each, with changes in test composition compensating for the decreased time.[27]

1980 test and associated changes
The inclusion of the "Strivers" Score study was implemented. This study was introduced by The Educational Testing Service, which administers the SAT, and has been conducting research on how to make it easier for minorities and individuals who suffer from social and economic barriers. The original "Strivers" project, which was in the research phase from 1980–1994, awarded special "Striver" status to test-takers who scored 200 points higher than expected for their race, gender and income level. The belief was that this would give minorities a better chance at being accepted in to a college of higher standard, i.e. an Ivy League school. In 1992, the Strivers Project was leaked to the public; as a result the Strivers Project was terminated in 1993. After Federal Courts heard arguments from the ACLU, NAACP and the Educational Testing Service, the courts ordered the study to alter its data collection process, stating that only the age, race and zip code could be used to determine the test-takers eligibility for "Strivers" points. These changes were introduced to the SAT effective in 1994.

1994 changes
In 1994 the verbal section received a dramatic change in focus. Among these changes were the removal of antonym questions, and an increased focus on passage reading. The mathematics section also saw a dramatic change in 1994, thanks in part to pressure from the National Council of Teachers of Mathematics. For the first time since 1935, the SAT asked some non-multiple choice questions, instead requiring students to supply the answers. 1994 also saw the introduction of calculators into the mathematics section for the first time in the test's history. The mathematics section introduced concepts of probability, slope, elementary statistics, counting problems, median and mode.[27] The average score on the 1994 modification of the SAT I was usually around 1000 (500 on the verbal, 500 on the math). The most selective schools in the United States (for example, those in the Ivy League) typically had SAT averages exceeding 1400 on the old test.

1995 re-centering (raising median score back to 500)
The test scoring was initially scaled to make 500 the mean score on each section with a standard deviation of 100.[29] As the test grew more popular and more students from less rigorous schools began taking the test, the average dropped to about 428 Verbal and 478 Math. The SAT was "recentered" in 1995, and the average "new" score became again close to 500. Scores awarded after 1994 and before October 2001 are officially reported with an "R" (e.g. 1260R) to reflect this change. Old scores may be recentered to compare to 1995 to present scores by using official College Board tables,[30] which in the middle ranges add about 70 points to Verbal and 20 or 30 points to Math. In other words, current students have a 100 (70 plus 30) point advantage over their parents.
1995 re-centering controversy

Certain educational organizations viewed the SAT re-centering initiative as an attempt to stave off international embarrassment in regards to continuously declining test scores, even among top students. As evidence, it was presented that the number of pupils who scored above 600 on the verbal portion of the test had fallen from a peak of 112,530 in 1972 to 73,080 in 1993, a 36% backslide, despite the fact that the total number of test-takers had risen over 500,000.\[31\]

2002 changes – Score Choice

In October 2002, the College Board dropped the Score Choice Option for SAT-II exams. Under this option, scores were not released to colleges until the student saw and approved of the score.\[32\] The College Board has since decided to re-implement Score Choice in the spring of 2009. It is described as optional, and it is not clear if the reports sent will indicate whether or not this student has opted-in or not. A number of highly selective colleges and universities, including Yale, the University of Pennsylvania, and Stanford, have announced they will require applicants to submit all scores. Stanford, however, only prohibits Score Choice for the traditional SAT.\[33\] Others, such as MIT and Harvard, have fully embraced Score Choice.

2005 changes

In 2005, the test was changed again, largely in response to criticism by the University of California system.\[34\] Because of issues concerning ambiguous questions, especially analogies, certain types of questions were eliminated (the analogies from the verbal and quantitative comparisons from the Math section). The test was made marginally harder, as a corrective to the rising number of perfect scores. A new writing section, with an essay, based on the former SAT II Writing Subject Test, was added,\[35\] in part to increase the chances of closing the opening gap between the highest and midrange scores. Other factors included the desire to test the writing ability of each student; hence the essay. The New SAT (known as the SAT Reasoning Test) was first offered on March 12, 2005, after the last administration of the "old" SAT in January 2005. The Mathematics section was expanded to cover three years of high school mathematics. The Verbal section's name was changed to the Critical Reading section.

2008 changes

In late 2008, a new variable came into play. Previously, applicants to most colleges were required to submit all scores, with some colleges that embraced Score Choice retaining the option of allowing their applicants not to have to submit all scores. However, in 2008, an initiative to make Score Choice universal had begun, with some opposition from colleges desiring to maintain score report practices. While students theoretically now have the choice to submit their best score (in theory one could send any score one wishes to send) to the college of their choice, some popular colleges and universities, such as Cornell, ask that students send all test scores.\[36\] This had led the College Board to display on their web site which colleges agree with or dislike Score Choice, with continued claims that students will still never have scores submitted against their will.\[37\] Regardless of whether a given college permits applicants to exercise Score Choice options, most colleges do not penalize students who report poor scores along with high ones; many universities, such as Columbia and Cornell, expressly promise to overlook those scores that may be undesirable to the student and/or to focus more on those scores that are most representative of the student's achievement and academic potential. College Board maintains a list of colleges and their respective score choice policies that is recent as of November 2011.\[38\]
Name changes and recentered scores
The name originally stood for "Scholastic Aptitude Test". But in 1990, because of uncertainty about the SAT's ability to function as an intelligence test, the name was changed to Scholastic Assessment Test. In 1993 the name was changed to SAT I: Reasoning Test (with the letters not standing for anything) to distinguish it from the SAT II: Subject Tests. In 2004, the roman numerals on both tests were dropped, and the SAT I was renamed the SAT Reasoning Test. The scoring categories are now the following: Critical Reading (comparable to some of the Verbal portions of the old SAT I), Mathematics, and Writing. The writing section now includes an essay, whose score is involved in computing the overall score for the Writing section, as well as grammar sections (also comparable to some Verbal portions of the previous SAT).

The test scoring was initially scaled to make 500 the mean score on each section with a standard deviation of 100. The SAT was "recentered" in 1995, and the average "new" score became again close to 500. Scores awarded after 1994 and before October 2001 are officially reported with an "R" (e.g. 1260R) to reflect this change. Old scores may be recentered to compare to 1995 to present scores by using official College Board tables, which in the middle ranges add about 70 points to Verbal and 20 or 30 points to Math.

Scoring problems of October 2005 tests
In March 2006, it was announced that a small percentage of the SATs taken in October 2005 had been scored incorrectly due to the test papers being moist and not scanning properly, and that some students had received erroneous scores. The College Board announced they would change the scores for the students who were given a lower score than they earned, but at this point many of those students had already applied to colleges using their original scores. The College Board decided not to change the scores for the students who were given a higher score than they earned. A lawsuit was filed in 2005 by about 4,400 students who received an incorrect low score on the SAT. The class-action suit was settled in August 2007 when The College Board and another company that administers the college-admissions test announced they would pay $2.85 million to over 4,000 students. Under the agreement each student can either elect to receive $275 or submit a claim for more money if he or she feels the damage was even greater. A similar scoring error occurred on a secondary school admission test in 2010-2011 when the ERB (Educational Records Bureau) announced after the admission process was over that an error had been made in the scoring of the tests of 2010 (17%) of the students who had taken the Independent School Entrance Examination for admission to private secondary schools for 2011. Commenting on the effect of the error on students' school applications in the New York Times, David Clune, President of the ERB stated "It is a lesson we all learn at some point — that life isn't fair."

The math-verbal achievement gap
In 2002, Richard Rothstein (education scholar and columnist) wrote in The New York Times that the U.S. math averages on the SAT & ACT continued its decade-long rise over national verbal averages on the tests.

Criticism
Cultural bias
For decades many critics have accused designers of the verbal SAT of cultural bias toward the white and wealthy. A famous example of this bias in the SAT I was the oarsman–regatta analogy question. The object of the question was to find the pair of terms that have the relationship most similar to the relationship between "runner" and "marathon". The correct answer was "oarsman" and "regatta". The choice of the correct answer presupposed students' familiarity with crew, a sport popular with the wealthy, and so upon their knowledge of its structure and terminology. Fifty-three percent (53%) of white students correctly answered the question, while only 22% of black
students also scored correctly. However, according to Murray and Herrnstein, the black-white gap is smaller in culture-loaded questions like this one than in questions that appear to be culturally neutral. Analogy questions have since been replaced by short reading passages.

**Stress**

The SAT exams have been regarded as a prominent cause for stress among aspiring students. Many students feel that creativity will drive the future, and that SATs do not measure creativity. SATs are regarded to not measure the full capabilities of certain individuals as well.

**Test score disparity by income**

Recent research has linked high family incomes to higher mean scores. Test score data from California has shown that test-takers with family incomes of less than $20,000 a year had a mean score of 1310 while test-takers with family incomes of over $200,000 had a mean score of 1715, a difference of 405 points. The estimates of correlation of SAT scores and household income range from 0.23 to 0.4 (explaining about 5-16% of the variation). One calculation has shown a 40-point average score increase for every additional $20,000 in income. There are conflicting opinions on the source of this correlation. Some think it is evidence of superior education and tutoring that is accessible to the more affluent adolescents. Still others propose it relates to wealthier families being exposed to a broader range of cultural ideas and experiences, because of travel and other means of wider exposure, and that "Cultural Literacy" can lead to enhancement of aptitude.

**Dropping SAT**

A growing number of colleges have responded to this criticism by joining the SAT optional movement. These colleges do not require the SAT for admission.

In a 2001 speech to the American Council on Education, Richard C. Atkinson, the president of the University of California, urged dropping the SAT Reasoning Test as a college admissions requirement:

"Anyone involved in education should be concerned about how overemphasis on the SAT is distorting educational priorities and practices, how the test is perceived by many as unfair, and how it can have a devastating impact on the self-esteem and aspirations of young students. There is widespread agreement that overemphasis on the SAT harms American education."

In response to threats by the University of California to drop the SAT as an admission requirement, the College Entrance Examination Board announced the restructuring of the SAT, to take effect in March 2005, as detailed above.

In the 1960's and 70's there was a movement to drop achievement scores. After a period of time, the countries, states and provinces that reintroduced them agreed that academic standards had dropped, students had studied less and had taken their studying less seriously. They reintroduced the tests after studies and research concluded that the high stakes tests produced benefits that out weighed the costs.

**MIT study**

In 2005, MIT Writing Director Les Perelman plotted essay length versus essay score on the new SAT from released essays and found a high correlation between them. After studying over 50 graded essays, he found that longer essays consistently produced higher scores. In fact, he argues that by simply gauging the length of an essay without reading it, the given score of an essay could likely be determined correctly over 90% of the time. He also discovered that several of these essays were full of factual errors; the College Board does not claim to grade for factual accuracy.

Perelman, along with the National Council of Teachers of English also criticized the 25-minute writing section of the test for damaging standards of writing teaching in the classroom. They say that writing teachers training their
students for the SAT will not focus on revision, depth, accuracy, but will instead produce long, formulaic, and wordy pieces.\textsuperscript{51} "You’re getting teachers to train students to be bad writers", concluded Perelman.\textsuperscript{52}

**Test preparation**

SAT preparation is a highly lucrative field.\textsuperscript{53} Many companies and organizations offer test preparation in the form of books, classes, online courses, and tutoring. Although the College Board maintains that the SAT is essentially uncoachable, some research suggests that tutoring courses result in an average increase of about 20 points on the math section and 10 points on the verbal section.\textsuperscript{54}

**References**

[16] The scoring categories are the following, Reading, Math, Writing, and Essay.
Further reading


External links
- Official SAT Test website (http://sat.collegeboard.org/register/)

ACT

ACT may refer to:

- Australian Capital Territory, the capital territory of the Commonwealth of Australia
- ACT (test), a college placement exam in the United States

In business:

- ACT mouthwash from Chattem
- Advanced Cell Technology, a biotech corporation involved in researching and developing cloning and stem cell technologies
- Ada Core Technologies, now AdaCore, a software company that develops an Ada compiler, tools and libraries
- Applied Computer Techniques, renamed Apricot Computers in 1985, a computer services company
- Aviation Composite Technology, an aircraft manufacturer in the Philippines
- ACT (NASDAQ), Automated Confirmation of Transactions, a trade reporting and clearing system
- Advance corporation tax, a tax collected in the United Kingdom until 1999
- ACT!, contact management software

Organizations:

- ACT Alberta, a Canadian coalition to raise awareness about human trafficking
- ACT Alliance, an alliance of churches for coordinating humanitarian assistance and development work.
- ACT New Zealand, a political party in New Zealand
- Academy of Clinical Thyroidologists
- Addenbrooke's Charitable Trust
- Association for Competitive Technology, a technology lobby group
- The Association of Commercial Television in Europe, a group which represents the business interests of the commercial television sector at the EU institutions
- Association of Cinematograph Technicians, now the Association of Cinematograph Television and Allied Technicians
• Association of Corporate Treasurers
• Action for Children's Television, an organization dedicated to improvements in children's television
• Advanced Concepts Team, a research team of the European Space Agency
• Alliance of Concerned Teachers, an alliance of teachers, educators and teacher organizations in the Philippines
• America Coming Together, a United States political group
• AIDS Committee of Toronto, sponsor of Fashion Cares
• Asian College of Technology, a private university in the Philippines
• Association for Citizenship Teaching, a professional subject association for teachers of citizenship
• United Education Institute or Advanced Career Training, a private post-secondary educational facility with locations in Georgia and Florida
• Australian College of Theology
• Amazon Conservation Team, a non-profit organization that works in partnership with indigenous people to protect the Amazon rainforest

In government:
• Allied Command Transformation, a NATO strategic command
• Advanced Concepts Team, a research team of the European Space Agency

In science and technology:
• ACT-R, a symbolic cognitive architecture
• ACT (audio format), file format
• Active voice or Actor role in interlinear glossing
• Apple Certified Trainer, one of the Apple certification programs
• Astrographic Catalog/Tycho, a star catalogue; see Hipparcos Catalogue
• Atacama Cosmology Telescope, a telescope in Chile

In medicine:
• Artemisinin based combination therapy, a treatment for malaria
• Assertive community treatment, a system for treating mental illness
• Acceptance and commitment therapy, a type of cognitive-behavioral therapy
• Activated clotting time

In entertainment:
• A.C.T, a progressive rock band from Sweden
• ACT Theatre, in Seattle, Washington
• American Conservatory Theater, in San Francisco, California

In sport:
• Australia national cricket team
• American Canadian Tour, a race organisation.

In transport:
• Ford ACT, a people mover, short for Automatically Controlled Transportation or Activity Center Transit
• Ascot railway station, a railway station in the United Kingdom with the National Rail code "ACT"
• Waco Regional Airport, an airport in Waco, Texas, United States, with the IATA airport code "ACT"
• Agency for Community Transit, Madison County Transit's paratransit system
A standardized test is a test that is administered and scored in a consistent, or "standard", manner. Standardized tests are designed in such a way that the questions, conditions for administering, scoring procedures, and interpretations are consistent[1] and are administered and scored in a predetermined, standard manner.[2]

Any test in which the same test is given in the same manner to all test takers is a standardized test. Standardized tests need not be high-stakes tests, time-limited tests, or multiple-choice tests. The opposite of a standardized test is a non-standardized test. Non-standardized testing gives significantly different tests to different test takers, or gives the same test under significantly different conditions (e.g., one group is permitted far less time to complete the test than the next group), or evaluates them differently (e.g., the same answer is counted right for one student, but wrong for another student).

Standardized tests are perceived as being more fair than non-standardized tests. The consistency also permits more reliable comparison of outcomes across all test takers.

History

China

The earliest evidence of standardized testing was in China,[3] where the imperial examinations covered the Six Arts which included music, archery and horsemanship, arithmetic, writing, and knowledge of the rituals and ceremonies of both public and private parts. Later, the studies (military strategies, civil law, revenue and taxation, agriculture and geography) were added to the testing. In this form, the examinations were institutionalized during the 6th century CE, under the Sui Dynasty.

Britain

Standardized testing was introduced into Europe in the early 19th century, modeled on the Chinese mandarin examinations,[4] through the advocacy of British colonial administrators, the most "persistent" of which was Britain's consul in Guangzhou, China, Thomas Taylor Meadows. [4] Meadows warned of the collapse of the British Empire if standardized testing was not implemented throughout the empire immediately.[4]

Prior to their adoption, standardized testing was not traditionally a part of Western pedagogy; based on the sceptical and open-ended tradition of debate inherited from Ancient Greece, Western academia favored non-standardized assessments using essays written by students. It is because of this that the first European implementation of standardized testing did not occur in Europe proper, but in British India. [5] Inspired by the Chinese use of standardized testing, in the early 19th century, British "company managers hired and promoted employees based on competitive examinations in order to prevent corruption and favoritism."[5] This practice of standardized testing was later adopted in the late 19th century by the British mainland. The parliamentary debates that ensued made many references to the "Chinese mandarin system."[4]
It was from Britain, that standardized testing spread, not only throughout the British Commonwealth, but to Europe and then America. Its spread was fueled by the Industrial Revolution. Given the large number of school students during and after the Industrial Revolution, when compulsory education laws increased student populations, open-ended assessment of all students decreased. Moreover, the lack of a standardized process introduces a substantial source of measurement error, as graders might show favoritism or might disagree with each other about the relative merits of different answers.

More recently, it has been shaped in part by the ease and low cost of grading of multiple-choice tests by computer. Grading essays by computer is more difficult, but is also done. In other instances, essays and other open-ended responses are graded according to a pre-determined assessment rubric by trained graders.

**United States**

Further information: List of standardized tests in the United States

The use of standardized testing in the United States is a 20th-century phenomenon with its origins in World War I and the Army Alpha and Beta tests developed by Robert Yerkes and colleagues.

In the United States, the need for the federal government to make meaningful comparisons across a highly de-centralized (locally controlled) public education system has also contributed to the debate about standardized testing, including the Elementary and Secondary Education Act of 1965 that required standardized testing in public schools. US Public Law 107-110, known as the No Child Left Behind Act of 2001 further ties public school funding to standardized testing.

**Design and scoring**

Standardized testing can be composed of multiple-choice, true-false, essay questions, authentic assessments, or nearly any other form of assessment. Multiple-choice and true-false items are often chosen because they can be given and scored inexpensively and quickly by scoring special answer sheets by computer or via computer-adaptive testing. Some standardized tests have short-answer or essay writing components that are assigned a score by independent evaluators who use rubrics (rules or guidelines) and benchmark papers (examples of papers for each possible score) to determine the grade to be given to a response. Most assessments, however, are not scored by people; people are used to score items that are not able to be scored easily by computer (i.e., essays). For example, the Graduate Record Exam is a computer-adaptive assessment that requires no scoring by people (except for the writing portion).

**Scoring issues**

Human scoring is often variable, which is why computer scoring is preferred when feasible. For example, some believe that poorly paid employees will score tests badly. Agreement between scorers can vary between 60 to 85 percent, depending on the test and the scoring session. Sometimes states pay to have two or more scorers read each paper; if their scores do not agree, then the paper is passed to additional scorers.

Open-ended components of tests are often only a small proportion of the test. Most commonly, a major test includes both human-scored and computer-scored sections.
Score

There are two types of standardized test score interpretations: a norm-referenced score interpretation or a criterion-referenced score interpretation. Norm-referenced score interpretations compare test-takers to a sample of peers. Criterion-referenced score interpretations compare test-takers to a criterion (a formal definition of content), regardless of the scores of other examinees. These may also be described as standards-based assessments as they are aligned with the standards-based education reform movement. Norm-referenced test score interpretations are associated with traditional education, which measures success by rank ordering students using a variety of metrics, including grades and test scores, while standards-based assessments are based on the belief that all students can succeed if they are assessed against standards which are required of all students regardless of ability or economic background.

Standards

The considerations of validity and reliability typically are viewed as essential elements for determining the quality of any standardized test. However, professional and practitioner associations frequently have placed these concerns within broader contexts when developing standards and making overall judgments about the quality of any standardized test as a whole within a given context.

Evaluation standards

In the field of evaluation, and in particular educational evaluation, the Joint Committee on Standards for Educational Evaluation has published three sets of standards for evaluations. The Personnel Evaluation Standards was published in 1988, The Program Evaluation Standards (2nd edition) was published in 1994, and The Student Evaluation Standards was published in 2003.

Each publication presents and elaborates a set of standards for use in a variety of educational settings. The standards provide guidelines for designing, implementing, assessing and improving the identified form of evaluation. Each of the standards has been placed in one of four fundamental categories to promote educational evaluations that are proper, useful, feasible, and accurate. In these sets of standards, validity and reliability considerations are covered under the accuracy topic. For example, the student accuracy standards help ensure that student evaluations will provide sound, accurate, and credible information about student learning and performance.

Testing standards

In the field of psychometrics, the Standards for Educational and Psychological Testing place standards about validity and reliability, along with errors of measurement and issues related to the accommodation of individuals with disabilities. The third and final major topic covers standards related to testing applications, credentialing, plus testing in program evaluation and public policy.

Advantages

One of the main advantages of standardized testing is that the results can be empirically documented; therefore, the test scores can be shown to have a relative degree of validity and reliability, as well as results which are generalizable and replicable. This is often contrasted with grades on a school transcript, which are assigned by individual teachers. It may be difficult to account for differences in educational culture across schools, difficulty of a given teacher's curriculum, differences in teaching style, and techniques and biases that affect grading. This makes standardized tests useful for admissions purposes in higher education, where a school is trying to compare students from across the nation or across the world.

Another advantage is aggregation. A well designed standardized test provides an assessment of an individual's mastery of a domain of knowledge or skill which at some level of aggregation will provide useful information. That
is, while individual assessments may not be accurate enough for practical purposes, the mean scores of classes, schools, branches of a company, or other groups may well provide useful information because of the reduction of error accomplished by increasing the sample size.

Standardized tests, which by definition give all test-takers the same test under the same (or reasonably equal) conditions, are also perceived as being more fair than assessments that use different questions or different conditions for students according to their race, socioeconomic status, or other considerations.

Disadvantages and criticism

"Standardized tests can't measure initiative, creativity, imagination, conceptual thinking, curiosity, effort, irony, judgment, commitment, nuance, good will, ethical reflection, or a host of other valuable dispositions and attributes. What they can measure and count are isolated skills, specific facts and function, content knowledge, the least interesting and least significant aspects of learning."

— Bill Ayers

Standardized tests are useful tools for assessing student achievement, and can be used to focus instruction on desired outcomes, such as reading and math skills. However, critics feel that overuse and misuse of these tests harms teaching and learning by narrowing the curriculum. According to the group FairTest, when standardized tests are the primary factor in accountability, schools use the tests to define curriculum and focus instruction. Critics say that "teaching to the test" disfavors higher-order learning. While it is possible to use a standardized test without letting its contents determine curriculum and instruction, frequently, what is not tested is not taught, and how the subject is tested often becomes a model for how to teach the subject.

Uncritical use of standardized test scores to evaluate teacher and school performance is inappropriate, because the students' scores are influenced by three things: what students learn in school, what students learn outside of school, and the students' innate intelligence. The school only has control over one of these three factors. Value-added modeling has been proposed to cope with this criticism by statistically controlling for innate ability and out-of-school contextual factors. In a value-added system of interpreting test scores, analysts estimate an expected score for each student, based on factors such as the student's own previous test scores, primary language, or socioeconomic status. The difference between the student's expected score and actual score is presumed to be due primarily to the teacher's efforts.

Supporters of standardized testing respond that these are not reasons to abandon standardized testing in favor of either non-standardized testing or of no assessment at all, but rather criticisms of poorly designed testing regimes. They argue that testing does and should focus educational resources on the most important aspects of education — imparting a pre-defined set of knowledge and skills — and that other aspects are either less important, or should be added to the testing scheme.

In Cathy Davidson's book, Now You See It, Davidson criticizes standardized tests. She describes our youth as "assembly line kids on an assembly line model," meaning the use of standardized test as a part of a one-size-fits-all educational model. She also criticizes the narrowness of skills being tested and labeling children without these skills as failures or as students with disabilities.
Scoring information loss

When tests are scored right-wrong, an important assumption has been made about learning. The number of right answers or the sum of item scores (where partial credit is given) is assumed to be the appropriate and sufficient measure of current performance status. In addition, a secondary assumption is made that there is no meaningful information in the wrong answers.

In the first place, a correct answer can be achieved using memorization without any profound understanding of the underlying content or conceptual structure of the problem posed. Second, when more than one step for solution is required, there are often a variety of approaches to answering that will lead to a correct result. The fact that the answer is correct does not indicate which of the several possible procedures were used. When the student supplies the answer (or shows the work) this information is readily available from the original documents.

Second, if the wrong answers were blind guesses, there would be no information to be found among these answers. On the other hand, if wrong answers reflect interpretation departures from the expected one, these answers should show an ordered relationship to whatever the overall test is measuring. This departure should be dependent upon the level of psycholinguistic maturity of the student choosing or giving the answer in the vernacular in which the test is written.

In this second case it should be possible to extract this order from the responses to the test items. Such extraction processes, the Rasch model for instance, are standard practice for item development among professionals. However, because the wrong answers are discarded during the scoring process, attempts to interpret these answers for the information they might contain is seldom undertaken.

Third, although topic-based subtest scores are sometimes provided, the more common practice is to report the total score or a rescaled version of it. This rescaling is intended to compare these scores to a standard of some sort. This further collapse of the test results systematically removes all the information about which particular items were missed.

Thus, scoring a test right–wrong loses 1) how students achieved their correct answers, 2) what led them astray towards unacceptable answers and 3) where within the body of the test this departure from expectation occurred.

This commentary suggests that the current scoring procedure conceals the dynamics of the test-taking process and obscures the capabilities of the students being assessed. Current scoring practice oversimplifies these data in the initial scoring step. The result of this procedural error is to obscure of the diagnostic information that could help teachers serve their students better. It further prevents those who are diligently preparing these tests from being able to observe the information that would otherwise have alerted them to the presence of this error.

A solution to this problem, known as Response Spectrum Evaluation (RSE), is currently being developed that appears to be capable of recovering all three of these forms of information loss, while still providing a numerical scale to establish current performance status and to track performance change.

This RSE approach provides an interpretation of the thinking processes behind every answer (both the right and the wrong ones) that tells teachers how they were thinking for every answer they provide. Among other findings, this chapter reports that the recoverable information explains between two and three times more of the test variability than considering only the right answers. This massive loss of information can be explained by the fact that the "wrong" answers are removed from the test information being collected during the scoring process and is no longer available to reveal the procedural error inherent in right-wrong scoring. The procedure bypasses the limitations produced by the linear dependencies inherent in test data.

Testing bias occurs when a test systematically favors one group over another, even though both groups are equal on the trait the test measures. Critics allege that test makers and facilitators tend to represent a middle class, white background. Critics claim that standardized testing match the values, habits, and language of the test makers. However, being that most tests come from a white, middle-class background, it is important to note that the highest scoring groups are not people of that background, but rather tend to come from Asian populations.
Not all tests are well-written, for example, containing multiple-choice questions with ambiguous answers, or poor coverage of the desired curriculum. Some standardized tests include essay questions, and some have criticized the effectiveness of the grading methods. Recently, partial computerized grading of essays has been introduced for some tests, which is even more controversial.\[24\]

**Educational decisions**

Test scores are in some cases used as a sole, mandatory, or primary criterion for admissions or certification. For example, some U.S. states require high school graduation examinations. Adequate scores on these exit exams are required for high school graduation. The General Educational Development test is often used as an alternative to a high school diploma.

Other applications include tracking (deciding whether a student should be enrolled in the "fast" or "slow" version of a course) and awarding scholarships. In the United States, many colleges and universities automatically translate scores on Advanced Placement tests into college credit, satisfaction of graduation requirements, or placement in more advanced courses. Generalized tests such as the SAT or GRE are more often used as one measure among several, when making admissions decisions. Some public institutions have cutoff scores for the SAT, GPA, or class rank, for creating classes of applicants to automatically accept or reject.

Heavy reliance on standardized tests for decision-making is often controversial, for the reasons noted above. Critics often propose emphasizing cumulative or even non-numerical measures, such as classroom grades or brief individual assessments (written in prose) from teachers. Supporters argue that test scores provide a clear-cut, objective standard that minimizes the potential for political influence or favoritism.

The National Academy of Sciences recommends that major educational decisions not be based solely on a test score.\[25\] The use of minimum cut-scores for entrance or graduation does not imply a single standard, since test scores are nearly always combined with other minimal criteria such as number of credits, prerequisite courses, attendance, etc. Test scores are often perceived as the "sole criteria" simply because they are the most difficult, or the fulfillment of other criteria is automatically assumed. One exception to this rule is the GED, which has allowed many people to have their skills recognized even though they did not meet traditional criteria.

**References**

[1] Sylvan Learning Glossary, retrieved online, source no longer available
[7] ETS website (http://www.ets.org/portal/site/ets/menuitem.1488512ecd5b8849a77b13bc3921509/?vgnextoid=302433c7f00c5010VgnVCM10000022f95190RCRD&vgnextchannel=7196ce3bf64f4010VgnVCM10000022f95190RCRD#Scoring_and_Reporting) about scoring the GRE.
[8] Houtz, Jolayne (August 27, 2000) "Temps spend just minutes to score state test A WASL math problem may take 20 seconds; an essay, 2 1/2 minutes" (http://archives.seattletimes.nwsource.com/cgi-bin/texis.cgi/web/vortex/display?slug=4039520&query=wasl+pearson+pay+hour). *Seattle Times* "In a matter of minutes, a $10-an-hour temp assigns a score to your child's test"
[10] Joint Committee on Standards for Educational Evaluation (http://www.wmich.edu/evalctr/jc/)
[22] "A Paradigm Shift in Test Scoring!" (http://www.better-schooling-systems.org)

Further reading

External links
- Joint Committee on Standards for Educational Evaluation (http://www.wmich.edu/evalctr/jc/)
- Standardized Testing in School (http://www.worknwoman.com/articles/st_testing2.shtml)
SAT
Test (assessment)

A test or examination is an assessment intended to measure a test-taker's knowledge, skill, aptitude, physical fitness, or classification in many other topics (e.g., beliefs). A test may be administered orally, on paper, on a computer, or in a confined area that requires a test taker to physically perform a set of skills. Tests vary in style, rigor and requirements. For example, in a closed book test, a test taker is often required to rely upon memory to respond to specific items whereas in an open book test, a test taker may use one or more supplementary tools such as a reference book or calculator when responding to an item. A test may be administered formally or informally. An example of an informal test would be a reading test administered by a parent to a child. An example of a formal test would be a final examination administered by a teacher in a classroom or an I.Q. test administered by a psychologist in a clinic. Formal testing often results in a grade or a test score.¹ A test score may be interpreted with regards to a norm or criterion, or occasionally both. The norm may be established independently, or by statistical analysis of a large number of participants.

A standardized test is any test that is administered and scored in a consistent manner to ensure legal defensibility.² Standardized tests are often used in education, professional certification, psychology (e.g., MMPI), the military, and many other fields.

A non-standardized test is usually flexible in scope and format, variable in difficulty and significance. Since these tests are usually developed by individual instructors, the format and difficulty of these tests may not be widely adopted or used by other instructors or
A non-standardized test may be used to determine the proficiency level of students, to motivate students to study, and to provide feedback to students. In some instances, a teacher may develop non-standardized tests that resemble standardized tests in scope, format, and difficulty for the purpose of preparing their students for an upcoming standardized test. Finally, the frequency and setting by which a non-standardized test is administered are highly variable and are usually constrained by the duration of the class period. A class instructor may, for example, administer a test on a weekly basis or just twice a semester. Depending on the policy of the instructor or institution, the duration of each test itself may last for only five minutes to an entire class period.

In contrast to non-standardized tests, standardized tests are widely used, fixed in terms of scope, difficulty and format, and are usually significant in consequences. Standardized tests are usually held on fixed dates as determined by the test developer, educational institution, or governing body, which may or may not be administered by the instructor, held within the classroom, or constrained by the classroom period. Although there is little variability between different copies of the same type of standardized test (e.g., SAT or GRE), there is variability between different types of standardized tests.

Any test with important consequences for the individual test taker is referred to as a high-stakes test. A test may be developed and administered by an instructor, a clinician, a governing body, or a test provider. In some instances, the developer of the test may not be directly responsible for its administration. For example, Educational Testing Service (ETS), a nonprofit educational testing and assessment organization, develops standardized tests such as the SAT but may not directly be involved in the administration or proctoring of these tests. As with the development and administration of educational tests, the format and level of difficulty of the tests themselves are highly variable and there is no general consensus or invariable standard for test formats and difficulty. Often, the format and difficulty of the test is dependent upon the educational philosophy of the instructor, subject matter, class size, policy of the educational institution, and requirements of accreditation or governing bodies. In general, tests developed and administered by individual instructors are non-standardized whereas tests developed by testing organizations are standardized.
History

Ancient China was the first country in the world that implemented a nationwide standardized test, which was called the imperial examination. The main purpose of this examination was to select for able candidates for specific governmental positions.[4] The imperial examination was established by the Sui Dynasty in 605 AD and was later abolished by the Qing Dynasty 1300 years later in 1905. England had adopted this examination system in 1806 to select specific candidates for positions in Her Majesty's Civil Service. This examination system was later applied to education and it started to influence other parts of the world as it became a prominent standard (e.g. regulations to prevent the markers from knowing the identity of candidates), of delivering standardized tests.

Influence of World Wars on Testing

Both World War I and World War II made many people realize the necessity of standardized testing and the benefits associated with these tests. One main reason people saw the benefits was from the Army Alpha and Army Beta tests, which were used during WWI to determine human abilities. Along side the Army Alpha, the Stanford-Binet Intelligence Scale "added momentum to the testing movement."[5] Soon after, colleges and industry began using tests to help in accepting and hiring people based on performance of the test.

Another reason more tests began to come forth was that people were realizing that the distance between secondary education and higher education was widening after WWII. In 1952, the first Advanced Placement (AP) test was administered to begin closing the gap between high schools and colleges. [6]

Modern day use of tests

Education

Some countries such as the United Kingdom and France require all their secondary school students to take a standardized test on individual subjects such as the General Certificate of Secondary Education (GCSE) (in England) and Baccalauréat respectively as a requirement for graduation.[7] These tests are used primarily to assess a student's proficiency in specific subjects such as mathematics, science, or literature. In contrasts, high school students in other countries such as the United States may not be required to take a standardized test to graduate. Moreover, students in these countries usually take standardized tests only to apply for a position in a university program and are typically given the option of taking different standardized tests such as the ACT or SAT, which are used primarily to measure a student's reasoning skill.[8][9] High school students in the United States may also take Advanced Placement tests on specific subjects to fulfill university-level credit. Depending on the policies of the test maker or country, administration of standardized tests may be done in a large hall, classroom, or testing center. A proctor or invigilator may also be present during the testing period to provide instructions, to answer questions, or to prevent cheating.

Grades or test scores from standardized test may also be used by universities to determine if a student applicant should be admitted into one of its academic or professional programs. For example, universities in the United Kingdom admit applicants into their undergraduate programs based primarily or solely on an applicant's grades on pre-university qualifications such as the GCE A-levels or Cambridge Pre-U.[10][11] In contrast, universities in the United States use an applicant's test score on the SAT or ACT as just one of their many admission criteria to determine if an applicant should be admitted into one of its undergraduate programs. The other criteria in this case may include the applicant's grades from high school, extracurricular activities, personal statement, and letters of recommendations.[12] Once admitted, undergraduate students in the United Kingdom or United States may be
required by their respective programs to take a comprehensive examination as a requirement for passing their courses or for graduating from their respective programs.

Standardized tests are sometimes used by certain countries to manage the quality of their educational institutions. For example, the No Child Left Behind Act in the United States requires individual states to develop assessments for students in certain grades. In practice, these assessments typically appear in the form of standardized tests. Test scores of students in specific grades of an educational institution are then used to determine the status of that educational institution, i.e., whether it should be allowed to continue to operate in the same way or to receive funding.

Finally, standardized tests are sometimes used to compare proficiencies of students from different institutions or countries. For example, the Organisation for Economic Co-operation and Development (OECD) uses Programme for International Student Assessment (PISA) to evaluate certain skills and knowledge of students from different participating countries.\[13\]

**Licensing and certification**

Standardized tests are sometimes used by certain governing bodies to determine if a test taker is allowed to practice a profession, to use a specific job title, or to claim competency in a specific set of skills. For example, a test taker who intends to become a lawyer is usually required by a governing body such a governmental bar licensing agency to pass a bar exam.

**Immigration and naturalization**

Standardized tests are also used in certain countries to regulate immigration. For example, intended immigrants to Australia are legally required to pass a citizenship test as part of that country's naturalization process.\[14\]

**Competitions**

Tests are sometimes used as a tool to select for participants that have potential to succeed in a competition such as a sporting event. For example, serious skaters who wish to participate in figure skating competitions in the United States must pass official U.S. Figure Skating tests just to qualify.\[15\]

**Group memberships**

Tests are sometimes used by a group to select for certain types of individuals to join the group. For example, Mensa International is a high I.Q. society that requires individuals to score at the 98th percentile or higher on a standardized, supervised IQ test.\[16\]
Types of tests

Written tests

Written tests are tests that are administered on paper or on a computer. A test taker who takes a written test could respond to specific items by writing or typing within a given space of the test or on a separate form or document.

In some tests, where knowledge of many constants or technical terms is required to effectively answer questions, like Chemistry or Biology - the test developer may allow every test taker to bring with them a cheat sheet.

A test developer's choice of which style or format to use when developing a written test is usually arbitrary given that there is no single invariant standard for testing. Be that as it may, certain test styles and format have become more widely used than others. Below is a list of those formats of test items that are widely used by educators and test developers to construct paper or computer-based tests. As a result, these tests may consist of only one type of test item format (e.g., multiple choice test, essay test) or may have a combination of different test item formats (e.g., a test that has multiple choice and essay items).

Multiple choice

In a test that has items formatted as multiple choice questions, a candidate would be given a number of set answers for each question, and the candidate must choose which answer or group of answers is correct. There are two families of multiple choice questions. The first family is known as the True/False question and it requires a test taker to choose all answers that are appropriate. The second family is known as One-Best-Answer question and it requires a test taker to answer only one from a list of answers. 

There are several reasons to using multiple choice questions in tests. In terms of administration, multiple choice questions usually requires less time for test takers to answer, are easy to score and grade, provide greater coverage of material, allows for a wide range of difficulty, and can easily diagnose a test taker's difficulty with certain concepts. As an educational tool, multiple choice items test many levels of learning as well as a test taker's ability to integrate information, and it provides feedback to the test taker about why distractors were wrong and why correct answers were right. Nevertheless, there are difficulties associated with the use of multiple choice questions. In administrative terms, multiple choice items that are effective usually take a great time to construct. As an educational tool, multiple choice items do not allow test takers to demonstrate knowledge beyond the choices provided and may even encourage guessing or approximation due to the presence of at least one correct answer. For instance a test taker might not work out explicitly that 6.14 * 7.95 = 48.813, but knowing that 6 * 8 = 48, they would choose an answer close to 48. Moreover, test takers may misinterpret these items and in the process, perceive these items to be tricky or picky. Finally, multiple choice items do not test a test taker's attitudes towards learning because correct responses can be easily faked.
Alternative Response

True/False questions present candidates with a binary choice - a statement is either true or false. This method presents problems, as depending on the number of questions, a significant number of candidates could get 100% just by guesswork, and should on average get 50%.

Matching Type

A matching item is an item that provides a defined term and requires a test taker is to match identifying characteristic to the correct term.\[19\]

Completion Type

A fill-in-the-blank item provides a test taker with identifying characteristics and requires the test taker to recall the correct term.\[19\] There are two types of fill-in-the-blank tests. The easier version provides a word bank of possible words that will fill in the blanks. For some exams all words in the word bank are exactly once. If a teacher wanted to create a test of medium difficulty, they would provide a test with a word bank, but some words may be used more than once and others not at all. The hardest variety of such a test is a fill-in-the-blank test in which no word bank is provided at all. This generally requires a higher level of understanding and memory than a multiple choice test. Because of this, fill-in-the-blank tests[with no word bank] are often feared by students.

Essay

Items such as short answer or essay typically require a test taker to write a response to fulfill the requirements of the item. In administrative terms, essay items take less time to construct.\[18\] As an assessment tool, essay items can test complex learning objectives as well as processes used to answer the question. The items can also provide a more realistic and generalizable task for test. Finally, these items make it difficult for test takers to guess the correct answers and require test takers to demonstrate their writing skills as well as correct spelling and grammar.

The difficulties with essay items is primarily administrative. For one, these items take more time for test takers to answer.\[18\] When these questions are answered, the answers themselves are usually poorly written because test takers may not have time to organize and proofread their answers. In turn, it takes more time to score or grade these items. When these items are being scored or graded, the grading process itself becomes subjective as non-test related information may influence the process. Thus, considerable effort is required to minimize the subjectivity of the grading process. Finally, as an assessment tool, essay questions may potentially be unreliable in assessing the entire content of a subject matter.

Mathematical questions

Most mathematics questions, or calculation questions from subjects such as chemistry, physics or economics employ a style which does not fall in to any of the above categories, although some papers, notably the Maths Challenge papers in the United Kingdom employ multiple choice. Instead, most mathematics questions state a mathematical problem or exercise that requires a student to write a freehand response. Marks are given more for the steps taken than for the correct answer. If the question has multiple parts, later parts may use answers from previous sections, and marks may be granted if an earlier incorrect answer was used but the correct method was followed, and an answer which is correct (given the incorrect input) is returned.

Higher level mathematical papers may include variations on true/false, where the candidate is given a statement and asked to verify its validity by direct proof or stating a counterexample.
Physical fitness tests

A physical fitness test is a test designed to measure physical strength, agility, and endurance. They are commonly employed in educational institutions as part of the physical education curriculum, in medicine as part of diagnostic testing, and as eligibility requirements in fields that focus on physical ability such as military or police. Throughout the 20th century, scientific evidence emerged demonstrating the usefulness of strength training and aerobic exercise in maintaining overall health, and more agencies began to incorporate standardized fitness testing. In the United States, the President's Council on Youth Fitness was established in 1956 as a way to encourage and monitor fitness in schoolchildren.

Common tests\cite{20}\cite{21}\cite{22} include timed running or the multi-stage fitness test, and numbers of push-ups, sit-ups/abdominal crunches and pull-ups that the individual can perform. More specialised tests may be used to test ability to perform a particular job or role.

Performance tests

A performance test is an assessment that requires an examinee to actually perform a task or activity, rather than simply answering questions referring to specific parts. The purpose is to ensure greater fidelity to what is being tested.

An example is a behind-the-wheel driving test to obtain a driver's license. Rather than only answering simple multiple-choice items regarding the driving of an automobile, a student is required to actually drive one while being evaluated.

Performance tests are commonly used in workplace and professional applications, such as professional certification and licensure. When used for personnel selection, the tests might be referred to as a work sample. A licensure example would be cosmetologists being required to demonstrate a haircut or manicure on a live person. The Group-Bourdon test is one of a number of psychometric tests which trainee train drivers in the UK are required to pass.\cite{23}

Some performance tests are simulations. For instance, the assessment to become certified as an ophthalmic technician includes two components, a multiple-choice examination and a computerized skill simulation. The examinee must demonstrate the ability to complete seven tasks commonly performed on the job, such as retinoscopy, that are simulated on a computer.

Test preparations

From the perspective of a test developer, there is great variability with respect to time and effort needed to prepare a test. Likewise, from the perspective of a test taker, there is also great variability with respect to the time and needed to obtain a desired grade or score on any given test. When a test developer constructs a test, the amount of time and effort is dependent upon the significance of the test itself, the proficiency of the test taker, the format of the test, class size, deadline of test, and experience of the test developer.

The process of test construction has been greatly aided in several ways. For one, many test developers were themselves students at one time, and therefore are able to modify or outright adopt test questions from their previous
tests. In some countries such as the United States, book publishers often provide teaching packages that include test banks to university instructors who adopt their published books for their courses. These test banks may contain up to four thousand sample test questions that have been peer-reviewed and time tested.[24] The instructor who chooses to use this test bank would only have to select a fixed number of test questions from this test bank to construct a test.

As with test constructions, the time needed for a test taker to prepare for a test is dependent upon the frequency of the test, the test developer, and the significance of the test. In general, nonstandardized tests that are short, frequent, and do not constitute a major portion of the test taker's overall course grade or score require do not require the test taker to spend great amounts preparing for the test.[25] Conversely, nonstandardized tests that are long, infrequent, and do constitute a major portion of the test taker's overall course grade or score usually require the test taker to spend great amounts preparing for the test. To prepare for a nonstandardized test, test takers may rely upon their reference books, class or lecture notes, Internet, and past experience to prepare for the test. Test takers may also use various learning aids to study for tests such as flash cards and mnemonics.[26] Test takers may even hire tutors to coach them through the process so that they may increase the probability of obtaining a desired test grade or score. Finally, test takers may rely upon past copies of a test from previous years or semesters to study for a future test. These past tests may be provided by a friend or a group that has copies of previous tests or from instructors and their institutions.[27]

Unlike nonstandardized test, the time needed by test takers to prepare for standardized tests are less variable and usually considerable. This is because standardized tests are usually uniformed in scope, format, and difficulty and often have important consequences with respect to a test taker's future such as a test taker's eligibility to attend a specific university program or to enter a desired profession. It is not unusual for test takers to prepare for standardized tests by relying upon commercially available books that provide in-depth coverage of the standardized test or compilations of previous tests (e.g., 10 year series in Singapore). In many countries, test takers even enroll in test preparation centers or cram schools that provide extensive or supplementary instructions to test takers to help them better prepare for a standardized test. Finally, in some countries, instructors and their institutions have also played a significant role in preparing test takers for a standardized test.

**Cheating on tests**

Cheating on a test is the process of using unauthorized means or methods for the purpose of obtaining a desired test score or grade. This may range from bringing and using notes during a closed book examination, copying another test taker's answer or choice of answers during an individual test, or even sending a paid proxy to take the test.

Several common methods have been employed to combat cheating. They include the use of multiple proctors or invigilators during a testing period to monitor test takers. Test developers may construct multiple variants of the same test to be administered to different test takers at the same time. In some cases, instructors themselves may not administer their own tests but will leave the task to other instructors or invigilators, which may mean that the invigilators do not know the candidates, and thus some form of identification may be required. Finally, instructors or test providers may compare the answers of suspected cheaters on the test themselves to determine if cheating did occur.
Support and criticisms of tests

Despite their widespread use, the validity, quality, or use of tests, particularly standardized tests in education have continued to be widely supported or criticized. Like the tests themselves, supports and criticisms of tests are often varied and may come from a variety of sources such as parents, test takers, instructors, business groups, universities, or governmental watchdogs.

Supporters of standardized tests in education often provide the following reasons for promoting testing in education:

- Feedback or diagnosis of test taker's performance[28]
- Fair and efficient[29]
- Promotes accountability[28][29]
- Prediction and selection[28]
- Improves performance[28]

Critics of standardized tests in education often provide the following reasons for revising or removing standardized tests in education:

- Narrows curricular format and encourages teaching to the test.[3]
- Poor predictive quality.[30]
- Grade inflation of test scores or grades[31][32][33]
- Culturally or socioeconomically biased.[34]

References

[2] North Central Regional Educational Laboratory, NCREL.org (http://www.ncrel.org/sdrs/areas/issues/students/earlycld/ea5lk3.htm)
[4] Advanced Level Examination, Chinese Language and Culture, Paper 1A
[13] "PISA" (http://www.pisa.oecd.org/pages/0,2987,en_32252351_32235731_1_1_1_1_1_1,00. html). .
[15] "Welcome to U.S. Figure Skating" (http://www.usfigureskating.org/Clubs.asp?id=353). .
[26] "Study strategies" (http://www.k-state.edu/counseling/topics/career/studystr.htm). .
Further reading


External links

Image Sources, Licenses and Contributors


