

Interpreting SAT and ACT Scores

John R. Hills
Florida State University

Notice This Material May Be Protected
By Copyright Law (Title 17 U.S. Code)

Millions of high school students each year take one of the major college admissions tests: the Scholastic Aptitude Test (SAT) of the College Board or the American College Test (ACT) of the American College Testing Program (ACTP). High school teachers and counselors assist students in deciding the meaning of the scores. Scales for the scores of these tests are different, and there is much confusion about what inferences can be derived. Often, those who are trying to make effective use of the scores are unaware of the mistakes they are making. Perhaps the following quiz and its explanations can help make these scores more useful in practice. Try these questions. The correct answers and brief explanations appear on page 50. This "test" may be reproduced without written permission. Please cite *Educational Measurement: Issues and Practice* as the source.

Hills' Handy Hints: SAT and ACT Scores

Please circle either Y or N, yes or no, for each question. Mark a response for each question as there is no penalty for guessing. This is an untimed test.

Y N 1. Donald told Mr. Henkin, the counselor at Burnside High School, that he had just received his Scholastic Aptitude Test (SAT) scores, and his Verbal (V) and Mathematical (M) scores were both about 400. He was disappointed because those scores are well below average. Is the average SAT score 500?

Y N 2. Mary told her best friend, Susie, that her SAT score was 900. Susie was impressed, because her scores were only V of 450 and M of 450. Did Mary perform better than Susie?

Y N 3. Jeremy received his SAT scores and his ACT scores in the mail. His ACT scores were much lower than his SAT scores. He had two SAT scores (495 and 515) but his ACT report had 5 scores, with numbers such as 17 and 18. Adding all 5 ACT scores together won't produce a total as high as one of the SAT scores. Did he perform less satisfactorily on the ACT?

Y N 4. Susie added together her V and M scores on the SAT and got a total of 1225. She looked at her ACT scores of 22, 23, 21, and 23, and expected them to add up to 89, but her ACT composite score is only 22. Has her ACT composite been calculated incorrectly?

Y N 5. Harold took both the SAT

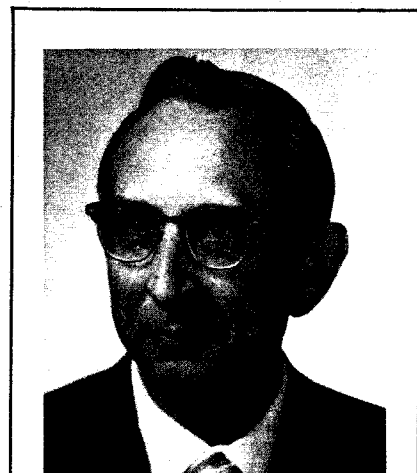
and the ACT. He estimated that for college-going students the SAT score mean is about 450 and the ACT score mean is about 17. However, he got ACT scores with an average near 20, but his SAT scores are only around 400. Is something wrong because he is above the mean on one test but below the mean on the other?

Y N 6. Tulawney, the star student of Burnside High School, took both the SAT and the ACT in the fall of her senior year. She scored near 700 on each of the SAT scores and near 25 on each of the ACT scores. She had so much fun taking these tests and did so well that at the next opportunity she took them both again. To her dismay, her scores on both the SAT and the ACT were lower on the second testing. She went to Mr. Henkin to find out what was wrong, and he told her not to worry. She should have expected her scores to be lower on the second testing. Is Mr. Henkin correct?

Y N 7. Willie took the SAT one year and got a Verbal score of 450. He was not satisfied, so he studied vocabulary diligently for a year and then took the test again. His Verbal score went up to 500. He recommended to everyone that they study vocabulary the way he did because his

score improved so much after only one year. Is Willie's vocabulary study responsible for his improved SAT Verbal score?

Y N 8. Sue Ellen wants to go to college very much, but her ACT composite score in her junior year was only 11. In order to raise that score, she studied diligently during the next year. She wanted to raise her achievement in natural science, social science, and mathematics because ACT tests include ques-



John Hills is Professor of Educational Measurement in the Department of Educational Research, Development and Foundations of the College of Education at Florida State University.

tions based on high school courses in those areas. When she took the ACT again, her composite went up to 14. The school principal, on hearing this, asked at a faculty meeting whether this kind of diligent study should be prescribed for all students who have low ACT scores. Mr. Henkin said that Sue Ellen's experience did not justify such a conclusion. Her improvement might readily have occurred without the extra study. Is that true?

Y N 9. Elijah took the Preliminary Scholastic Aptitude Test (PSAT) in his junior year and received a V score of 65 and an M score of 63. He asked Donald to help him decide what these scores mean. Donald said that they were very low. Donald was disappointed in his SAT scores, which were about 400,

but that is much higher than scores of only 65 and 63. Then Elijah went to Sue Ellen for advice. She had been thrilled to get her ACT score composite up to 14 after a year of hard work. She said that scores in the sixties were very high. Elijah had been influenced by a statistician, so he concluded that if one person said his scores were low and the other said they were high, the scores must be about average. Is that correct?

Y N 10. Mr. Livingston, the Commissioner of Education, is working hard to improve education in his state. He looked at the mean SAT scores for students in his state this year and found them little different from last year. Have his efforts been fruitless because the mean SAT scores did not increase?

Continued on page 50

Answers

1. N, 2. N, 3. N, 4. N, 5. N, 6. Y, 7. N,
8. Y, 9. N, 10. N

Explanations:

1. The "average" SAT score depends on whose data are being averaged. The score 500 was the average of 10,654 students who took the SAT in April of 1941, the first year that the aptitude tests of the College Entrance Examination Board (now the College Board) were given in objective-test format. (At that time, the tests were given only once a year, in April.) Since then, a process called equating has been used to insure that the aptitude level represented by 500 in April, 1941, is the same level represented by a score of 500 on an SAT taken at any other time. However, in 1983, the mean scores of all college-bound seniors tested were 430 on Verbal and 420 on Mathematics, for boys, and 493 and 445, respectively, for girls. The means for all high school seniors would, of course, be appreciably lower. We know now that mean SAT scores fluctuate from year to year, and until recently they have been declining. The score 500 is probably not the mean for any existing naturally-constituted group.

2. Mary must have added together her SAT-V and SAT-M scores to get 900. The highest reported score for either test is 800, so 900 cannot be a score from either section by itself. Susie's scores added together are 900, the same as Mary's scores. The College Board does not advocate adding V and M scores together for any purpose. In fact, if they are to be combined for estimating future success in college, experience indicates that a better general-purpose combination would be $2V+M$, with a minimum of 600 and a maximum of 2400.

3. The score scales used to report SAT and ACT results are dramatically different. The SAT scale is equated to a reference population that was arbitrarily assigned a mean of 500 and a standard deviation of 100. The ACT is equated to a different reference population that was arbitrarily assigned a low score of 1 and a high score of 36, making the mean about 18 and the standard deviation about 6. The test content is so different that the tests are far from parallel or equivalent, and their scores are not interchangeable

and would not be interchangeable even if calibrated to the same score scale. In fact, there is no psychometrically sound way to convert scores from one test to the other. There is no reason to expect the scores to agree with each other except roughly. The scales and content are so different that Jeremy is best advised that it is nonsense to try to determine from the scaled scores whether he did better on one test than the other.

4. Susie's ACT composite score of 22 is the average of her scaled scores on the four parts of the test. That is the way ACTP calculates the composite. Traditionally, SAT scores are kept separate and never averaged, whereas ACT scores are regularly averaged to form a composite score for the entire battery.

5. Scores from the two test batteries cannot be compared this way for several reasons. The principal reason is that the reference populations are different, and the groups of students taking the tests are probably different. They come from different regions of the United States, with different mixes of public and private colleges and different traditions of selectivity in college admissions. The college-bound students of the College Board colleges may be different kinds of students than the college-bound students of colleges served by ACTP. Beyond that, the contents of the two batteries of tests are markedly different. The SAT concentrates on items that are not tied to any high school subjects but emphasize developed ability that grows slowly through years of schooling. The ACT battery includes tests that heavily emphasize the skill of reading. Further, ACT tests ask questions that depend on specific knowledge students should have learned in high school mathematics, social studies, and natural sciences classes. Considering the differences in reference groups and test content, there is no reason to expect close similarity between an individual's specific scores on the two tests.

6. Tulawney achieved high SAT scores on her first testing, on the order of three standard deviations above the mean of a national sample of high school seniors. Because of imperfect reliability of the tests, which have reliability coefficients around .90, we would expect regression to the mean on a retest with a different form after a short time period allowing little growth.

The mean is lower than Tulawney's score, so we would expect that on the average, people who scored at Tulawney's level would obtain lower scores on retest.

7. Willie's improvement from 1 year to the next of 50 points on the SAT verbal could result from several factors outside of any study of vocabulary. Because his first score was near the mean, we would not expect much effect from statistical regression due to imperfect correlation between the two forms of the test. However, we do expect normal growth in developed abilities during a year. The standard error of measurement for an SAT scaled score is about 30 points. That means that about 68% of the time, people who receive an SAT score of a certain value on one testing will receive SAT scores within 30 points of that value on another testing, assuming there is no growth or learning in the interval. The combination of expected growth and variation due to error of measurement commonly produce increases of 50 points, without any special study or coaching.

8. Sue Ellen achieved a low composite on her first test, about one standard deviation below the mean for college-bound students who take ACT. Thus we should expect some regression to the mean upon retest. In addition to that, we would expect some increase in score due to normal growth and development. Finally, the standard error of measurement on the ACT composite score is one point, and we would expect the scores of some students to vary by as much as 2 standard errors of measurement from one testing to another. Those three factors could account for 3 score points improvement in the composite without attributing any of the change to special study or coaching.

(Note: This discussion does not mean to imply that study is a useless activity! Instead, it implies that little can be inferred from the experiences of one or two students, without adequate experimental controls, in the face of various statistical phenomena such as regression and errors of measurement.)

9. Elijah is not in the middle. The PSAT score scale is different from the scales of both the ACT and the SAT. Actually, the PSAT was designed to allow students to take a low-cost practice test for the SAT. PSAT forms are "retired" forms of the SAT: retired in

the sense that they were once used as regular SAT forms and will not ever be used again for that purpose. PSAT scores are highly correlated with SAT scores, but the scores are not interchangeable. To keep them separate, the last digit of the SAT score's 3-digit scale is left off. Scores of 65 and 63 would roughly correspond to SAT scores of 650 and 630, which are very good scores.

10. Mr. Livingston is making several mistakes when he looks at the mean SAT scores in his state to evaluate the efforts of the schools. First, those who take the SAT are not a representative sample of students in

the schools. They are self-selected, and temporary influences can dramatically change from year to year who selects himself or herself to consider going to a college that requires an admission test. In some states, the quality of the crops determines who will have the money for college. In others, it may be the sales of domestic automobiles. Beyond that, other factors determine whether students will take ACT admissions tests, SAT tests, or both. To complicate matters further, SAT tests are relatively insensitive to any short-term changes in schooling or study. Scores can be increased by better education, but the amount of increase depends on

the duration of the improved education, and short-term improvements that are of small scale (involving only modest changes or short periods of study each day or week) have little effect. Unless Mr. Livingston has made some sweeping and dramatic changes that have lasted several years, and unless he studies mean SAT scores of representative samples of students over time periods as long as a decade, he is unlikely to be able to observe the effects of his efforts. (Frank Womer discussed a similar common error in interpreting mean SAT scores for different states in *EM*, Vol. 2, No. 2, pp. 4-5.) ■