DOCUMENT RESUME

ED 383 715 TM 023 136

AUTHOR Stricker, Lawrence J.
TITLE SAT Scores and Academic Performance in High School.
INSTITUTION Educational Testing Service, Princeton, N.J.
REPORT NO ETS-RR-90-1
PUB DATE Jan 90
NOTE 22p.
PUB TYPE Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Academic Achievement; *Class Rank; *College Entrance Examinations; Correlation; *Grade Point Average; High Schools; *High School Students; Prediction; Scores; Test Construction; *Test Format; Test Results; Test Validity
IDENTIFIERS *Scholastic Aptitude Test

ABSTRACT

The aim of this study was to appraise whether different forms of the Scholastic Aptitude Test (SAT) used since the mid 1970s varied in their correlations with academic performance criteria in the same cohort of examinees. A 1975 form and a 1985 form were administered to equivalent samples of high school juniors, and self-reported grade-point average and high school rank were obtained. The SAT Verbal and Mathematical scores generally had similar correlations with the grade criteria in the two samples, but the Verbal score had a significantly higher correlation with school rank in the 1985 sample. The principal conclusion is that the 1975 form of the SAT does not have greater validity than the 1985 form in assessing academic performance, at least at the high school level. This outcome offers no support for the hypothesis that the decline in the SAT's ability to predict college grades since the mid 1970s, observed in recent research, is attributable to changes in the test. Three tables present analysis details. (Contains seven references.) (Author)

RESEARCH REPORT

SAT SCORES AND ACADEMIC PERFORMANCE IN HIGH SCHOOL

Lawrence J. Stricker

Educational Testing Service
Princeton, New Jersey
January 1990

SAT Scores and Academic Performance in High School

Lawrence J. Stricker

Research Report

January, 1990

Abstract

The aim of this study was to appraise whether different forms of the SAT used since the mid 1970s varied in their correlations with academic performance criteria in the same cohort of examinees. A 1975 form and a 1985 form were administered to equivalent samples of high school juniors, and self-reported grade-point average and high school rank were obtained. The SAT Verbal and Mathematical scores generally had similar correlations with the grade criteria in the two samples, but the Verbal score had a significantly higher correlation with school rank in the 1985 sample. The principal conclusion is that the 1975 form of the SAT does not have greater validity than the 1985 form in assessing academic performance, at least at the high school level. This outcome offers no support for the hypothesis that the decline in the SAT's ability to predict college grades since the mid 1970s, observed in recent research, is attributable to changes in the test.

SAT Scores and Academic Performance in High School

Recent research by Ramist (1984) and Morgan (1989) has documented systematic trends in the SAT's validity in predicting freshman grade-point average (GPA) in college. This work, using data from the Validity Study Service (VSS) on the correlations between the SAT and GPA within colleges participating in that service, found that the validity increased in the early 1970s and decreased since then.

The interpretation of those trends is complicated by a variety of changes that may have occurred during that same period of time. Obvious possibilities include changes in the samples of schools participating in the VSS, their student bodies, the SAT and its use in college admission, and grading practices (Morgan, 1989). The Morgan findings indicate that statistical analyses which take into account changes in the samples of schools and their student bodies reduce, but do not eliminate the decline in validity from the mid 1970s.

The present study addresses the possibility that changes in the SAT, whether intended or not, may be implicated in the decline from the mid 1970s. Some explicit changes have taken place in the test content and the test construction procedures, but it is uncertain that they are sufficient to affect predictive validity (see the review by Marco, Crone, Braswell, Curley, & Wright, in press). All the changes appear to be relatively minor, as can be judged from the following summary of the most notable ones. Starting in 1977, one reading passage for the reading comprehension items has had content of special relevance to members of minority groups. In 1978, the five long reading passages for the reading comprehension items were replaced by three long and three medium-length passages, the additional passage having science content. In 1982, the number of medium-difficulty items in the Verbal section

were increased, and the number of very difficult and very easy items were reduced. Finally, sensitivity reviews of the items to eliminate content that might be offensive to women and members of minority groups began on a routine basis in 1980.

The specific purpose of this study was to assess whether different forms of the SAT used since the mid 1970s varied in their correlations with academic performance criteria in the same cohort of examinees. Using the same cohort eliminates the effects of variations in schools, student bodies, and grading practices that complicate interpretation of both the Ramist and the Morgan studies of different cohorts. The criteria in question concern high school performance. The SAT is designed to predict college performance, but the SAT's relationships with performance in other academic settings are also relevant from the standpoint of the test's construct validity. Some evidence already exists about the SAT's associations with high school performance. Most pertinent is the research by N. W. Burton (personal communication, September 15, 1989), which found substantial correlations (_rs in the .40s and .50s) between the SAT and self-reported GPA and high school rank for 1975, 1980, and 1985 College Bound Seniors as well as a progressive decline in the correlations with GPA but not high school rank. Of course, the use of different cohorts complicates interpretation of the trend data.

Method

Data were drawn from a larger study, in which an SAT form initially used in 1975 (XSA2) and an SAT form just employed in 1985 (3HSA02) were administered in Spring 1989 to juniors in 862 high schools. The two forms constituted two of thirty spirals of actual and experimental SAT sections given with standard SAT test administration procedures. The schools and the

students volunteered to take part in the study (the latter were promised feedback about their test results). Self-reports of GPA and school rank, as well as other student characteristics, were also collected during the testing.[1] Data on characteristics of the schools were obtained from the AI Profile Survey conducted in 1988.

Usable data were obtained for 1,554 students in the 1975 sample and 1,753 in the 1985 sample, after excluding 427 students and 460 students from the respective samples. A total of 147 students in the 1975 sample and 151 in the 1985 sample were excluded because their schools departed from standard test administration procedures or participated in a pretest for the study, possibly exposing the students to the same tests; the students were not juniors planning to attend college within two years; or the students attempted three or fewer items in a test section. An additional 280 students in the 1975 sample and 309 in the 1985 sample were excluded because they did not report, in response to a question about their level of effort on the test, that they tried to do their best on "all or most of the questions" or "those questions that seemed easy or moderately difficult" (i.e., they chose a response indicative of a lower level of effort or did not respond at all).[2] Outlier analyses were also carried out with the remaining students to identify students with discrepant formula scores on the corresponding Verbal and Mathematical sections. No students with outlying scores were identified, using a test of discordancy for multivariate normal samples (Barnett & Lewis, 1985), with a familywise significance level of .25.

The characteristics of the samples are summarized in Table 1. The two samples were highly similar, both with regard to the characteristics of the students themselves and their schools.

Insert Table 1 about here

For each SAT form, converted Verbal and Mathematical scores, as well as formula scores for each item type (Antonyms, Analogies, Sentence Completion, Reading Comprehension, Regular Mathematics, Quantitative Comparisons) were obtained.

GPA was scaled from 4.3 (A+) to 0 (E/F). School rank was scaled from 4 (top 1% to 3%) to 1 (bottom third), each score representing a standard deviation interval (+1.5 to +2.5 to -.5 to -1.5) for the normal deviates corresponding to the percentile ranks.[3] School rank was excluded for students from schools with fewer than 25 students in the junior class because of the uncertain applicability of this variable in small classes. GPA and school rank were also combined in a grade composite, standardizing the GPA and school rank scores and computing their mean. (When one of the scores was missing, the grade composite was based on the available score.)

For each sample, product-moment correlations of the Verbal and Mathematical scores with the grade criteria were computed, the Ns varying because of missing data. The correlations of the item type scores with the criteria were also computed for exploratory purposes.

Results and Discussion

The means and standard deviations of the SAT scores and the grade criteria are reported in Table 2 for each sample. The data for the two samples were generally similar, the only marked differences occurring for some item type scores. The 1985 sample had somewhat lower Antonyms and

Quantitative Comparisons scores and somewhat higher Analogies and Reading Passages scores.

Insert Table 2 about here

It is noteworthy that the SAT data resemble those obtained when the test forms were originally administered, taking into consideration that the examinees in the original administrations included seniors as well as juniors (Stern & Bullock, 1976; N. Wright, D. Wright, & Weber, 1986). The SAT-V and SAT-M means for the 1975 sample in this study were 410 and 455, while the original means were 434 and 473; the means for the 1985 sample in this study were 405 and 453, and the original means were 423 and 470. This similarity suggests that the students in the study were adequately motivated, even though they were taking the test for research, and that the present results are generalizable to the regular SAT test-taking population.

GPA, school rank, and the grade composite had generally similar correlations in the two samples. GPA and school rank correlated .71 in the 1975 sample and .66 in the 1985 sample. Moreover, GPA correlated .95 with the grade composite in botn samples, and school rank correlated .92 and .93 with the grade composite in the two samples.

The correlations of the SAT scores with the grade criteria are reported in Table 3 for each sample. The SAT scores had generally similar correlations with GPA in the two samples (the sole exception was the Analogies item type score, which was significantly, $p < .05$, higher in the 1985 sample).[4] But SAT-V and two item type scores (Antonyms and Reading Comprehension) had significantly ($p < .05$) higher correlations with school rank in the 1985

sample. The SAT-V correlations with school rank were .39 in the 1975 sample and .46 in the 1985 sample. Two item type scores (Analogies and Quantitative Comparisons) also had significantly ($p < .05$) higher correlations with the grade composite in the 1985 sample.

Insert Table 3 about here

The principal conclusion of this study is that the 1975 form of the SAT does not have greater validity than the 1985 form in assessing academic performance, at least at the high school level. It is apparent that the outcome of this investigation offers no support for the hypothesis that the observed decline in the SAT's ability to predict college grades since the mid 1970s is attributable to changes in the test.

The differences in the pattern of correlations of the SAT Verbal score with GPA and school rank--similar correlations with GPA in the two samples but higher correlations with school rank in the 1985 sample--cannot be readily explained. The variances for school rank were similar in the two samples, and the samples do not appear to differ on other potentially relevant variables. It is noteworthy that the depressed correlation of the Verbal score with school rank in the 1975 sample (relative to the corresponding correlation with GPA in that sample), which contributed to the 1975 vs. 1985 difference, is mirrored in the Burton study. The Verbal score's correlations with GPA, for 1975 College Bound Seniors, were .50 for males and .51 for females; the corresponding correlations with school rank were .45 and .48.

It is important to emphasize that the results of this study concern the comparative validity of the two SAT forms with respect to academic performance

in today's high schools. The findings do not necessarily reflect what the validity of these forms would have been at an earlier point. Depending on the grading practices in previous years and the match between the test content and the then current curriculum, the validity of the forms could by very different. The Burton research is pertinent in this regard. The SAT's correlations with GPA and school rank for the 1975 and 1985 College Bound Seniors were generally similar to those in this study, with the important exception of the Verbal score's correlations with both grade criteria in the 1975 cohort. These correlations were substantially higher: for GPA, .50 (men) to .51 (women) vs. .43 in the present study; for school rank, .45 (men) to .48 (women) vs. .39 in this investigation.

Similarly, the present findings about the comparative validity of the SAT forms today do not bear on trends over time in the overlap of the SAT with high school GPA and school rank. Even if the 1985 form were unequivocally shown to have greater validity than the 1975 form against criteria of high school performance in 1989, it does not follow that the 1985 form is also more redundant with contemporary high school record variables or has less incremental validity vis-a-vis them.

Finally, the reliance on self-reported grades in this study inevitably raises a question about their usefulness for this purpose. It is well-established that self-reported high school grades are relatively accurate and correlate appreciably with recorded college grades (see the review by Baird, 1976), though not as highly as actual school grades (e.g., Morgan, 1989). The accuracy of the grades in this study is also supported by their substantial correlations with the SAT scores.

References

Baird, L. L. (1976). Using self-reports to predict student performance. New York: College Entrance Examination Board.

Barnett, V., & Lewis, T. (1984). Outliers in statistical data (2nd ed.). New York: Wiley.

Marco, G. L., Crone, C. R., Braswell, J. S., Curley, W. E., & Wright, N. K. (in press). Trends in SAT content and statistical characteristics and their relationship to SAT predictive validity (ETS Research Report). Princeton, NJ: Educational Testing Service.

Morgan, R. (1989). Analyses of the predictive validity of the SAT and high school grades from 1976 to 1985 (ETS Research Report 89-37; College Board Report 89-7). Princeton, NJ: Educational Testing Service.

Ramist, L. (1984). Predictive validity of the ATP tests. In T. F. Donlon (Ed.), The College Board technical handbook for the Scholastic Aptitude Test and Achievements Tests (pp. 141-170). New York: College Board.

Stern, J., & Bullock R. (1976). Test analysis--College Entrance Examination Board April 1975 Scholastic Aptitude Test, XSA2 (ETS Statistical Report 76-12). Princeton, NJ: Educational Testing Service.

Wright, D., Wright, N., & Weber, C. (1986). Test analysis--College Board Scholastic Aptitude Test March 1985 administration, 3HSA02 (ETS Statistical Report 86-11). Princeton, NJ: Educational Testing Service.

Author Notes

Thanks are due James S. Braswell, Elizabeth Burton, Sydell T. Carlton, Neil J. Dorans, Nancy J. Feryok, Carl H. Haag, Anne S. Neeff, and Nancy K. Wright for providing information and data about the Admissions Testing Program and the New Possibilities Project; William H. Angoff, Nancy W. Burton, Daniel R. Eignor, Charles Lewis, Samuel Messick, Donald A. Rock, and Warren W. Willingham for advising on the statistical analysis; Thomas J. Jirele for computer programming; and Nancy W. Burton, Charles Lewis, and Howard C. Wainer for reviewing a draft of this report.

Footnotes

[1]The GPA and school rank questions follow:

Give or estimate your grade point average.

A+

A

A-

B+

B

B-

C+

C

C-

D+

D

D-

E/F

In class rank, are you in the top, middle, or bottom third of your class?

Top third

Middle third

Bottom third

I don't know

If you named the top third to the previous question, about where do you stand in the top third?

1-3%

4-10%

11-20%

21-33%

[2]The question follows:

Which of the following best describes your efforts on this test? (Mark only one.)

I tried to do my best on all or most of the questions on this test.

I tried to do my best on those questions that seemed easy or moderately difficult.

I tried to do my best on those questions that were in the beginning of each section.

I tried to do my best on the first sections of the test.

I did not try to do my best on this test.

[3]The school rank categories were quantified as follows: 4 (top "1 to 3%"; 2.05 Z), 3 (top "4 to 10%," "11 to 20%," "21 to 33%"; 1.48 Z to .61 Z), 2 ("middle third"; .00 Z), and 1 ("bottom third"; -.97 Z).

[4]These two-tail significance tests are approximations because of the clustering within and between samples, with more than one student from each school in both samples.

Table 1

Characteristics of 1975 and 1985 Samples

Variable	1975		1985	
	N[b]	Mean or Percent	N[b]	Mean or Percent
Mean age	1525	16.59	1720	16.63
Percent female	1549	55.20	1743	55.82
Percent White	1552	73.90	1774	75.34
Percent with English as best language	1547	89.01	1740	88.91
Mean father's education (years)[a]	1313	14.25	1474	14.22
Mean mother's education (years)[a]	1359	13.84	1510	13.81
Percent in public school	1552	60.31	1751	63.16
Percent in school in large or medium-size city	1552	29.38	1751	28.56
Mean size of senior class in school	1411	178.55	1603	181.37
Percent in school offering college work	1552	37.44	1751	35.47
Percent in school offering honors courses	1552	29.38	1751	28.81

Note. For the 1975 sample, the standard deviations are .58 for Age, 3.28 for Father's Education, 2.99 for Mother's Education, and 165.02 for Size of Senior Class in School; for the 1985 sample, the corresponding standard deviations are .57, 3.36, 3.03, and 161.35.

[a] Education was quantified as follows: 18 (graduate or professional school), 16 (college graduate), 14 (some college), 12 (high school graduate or equivalent), and 6 (elementary school or some high school).

[b] Ns vary because of missing data.

Table 2

Means and Standard Deviations for SAT Scores and Grade Criteria in 1975 and 1985 Samples

Variable	1975		1985	
	Mean	S.D.	Mean	S.D.
SAT Total Score				
Verbal	410.37	103.14	405.37	106.08
Mathematical	454.91	113.07	453.15	121.32
SAT Item Type Score				
Antonyms	11.22	4.67	9.61	5.17
Analogies	6.85	4.24	8.25	4.43
Sentence Completion	5.68	3.35	6.09	2.94
Reading Comprehension	7.47	4.99	8.71	5.61
Regular Mathematics	14.01	8.85	14.82	8.42
Quantitative Comparisons	7.91	4.86	7.07	5.63
Grade Criterion				
Grade-Point Average	3.07	.68	3.08	.68
High School Rank	2.58	.68	2.63	.73
Grade Composite	-.03	.96	-.05	.95

Note. Verbal and Mathematical are converted scores; item types are formula scores. For the 1975 sample, the Ns are 1,544 for the SAT scores, 1,495 for Grade-Point Average, 947 for School Rank, and 1,525 for the Grade Composite; for the 1985 sample, the corresponding Ns are 1,753, 1,688, 1,100, and 1,720.

Table 3

Correlations of SAT Scores with Grade Criteria in 1975 and 1985 Sample

Variable	Grade-Point Average		School Rank		Grade Composite	
	1975	1985	1975	1985	1975	1985
SAT Total Score						
Verbal	.43	.45	.39	.46	.43	.48
Mathematical	.52	.51	.49	.50	.53	.54
SAT Item Type Score						
Antonyms	.36	.37	.31	.40	.35	.39
Analogies	.34	.40	.35	.40	.35	.42
Sentence Completion	.37	.39	.33	.38	.38	.40
Reading Comprehension	.41	.43	.36	.44	.42	.45
Regular Mathematics	.53	.49	.48	.49	.53	.52
Quantitative Comparisons	.44	.48	.42	.47	.45	.51

Note. Verbal and Mathematical are converted scores; item types are formula scores. For the 1975 sample, the Ns are 1,495 for the Grade-Point Average correlations, 947 for the School Rank correlations, and 1,525 for the Grade Composite correlations; for the 1985 sample, the corresponding Ns are 1,688, 1,100 and 1,720.