DOCUMENT RESUME

ED 217 065 TM 820 282

AUTHOR Alderman, Donald L.
TITLE Student Self-Selection and Test Repetition.
INSTITUTION Educational Testing Service, Princeton, N.J.
SPONS AGENCY College Entrance Examination Board, New York, N.Y.
REPORT NO CEEB-R-81-5; ETS-RR-81-22
PUB DATE 81
NOTE 15p.
AVAILABLE FROM College Board Publication Orders, Box 2815, Princeton, NJ 08541 ($4.00).

EDRS PRICE MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS *Comparative Analysis; Error of Measurement; High Schools; *High School Students; *Scores; Statistical Analysis; Testing Problems
IDENTIFIERS *Scholastic Aptitude Test; *Test Repeaters

ABSTRACT

The test performance of students who took the Scholastic Aptitude Test (SAT) only once as juniors was contrasted with students who took the test as juniors and again as seniors. Estimates of expected test performance on a common initial administration in the junior year were derived from separate equating sections and background variables. Residuals of observed minus expected test scores revealed statistically significant differences between students who took a single administration of the SAT as juniors and those who took the same initial administration but also repeated the test as seniors. The initial observed scores of students later repeating the test were consistently lower than their expected scores for both the verbal and mathematical sections. The results indicate that self-selection occurs when students decide to repeat a test. Score changes among these students reflect negative errors of measurement on the initial test administration. (Author/DWH)

College Board Report

No. 81-5

Student Self-Selection and Test Repetition

Donald L. Alderman

Student Self-Selection and Test Repetition

Donald L. Alderman

Educational Testing Service

College Board Report No. 81-5

ETS RR No. 81-22

College Entrance Examination Board, New York, 1981

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

Portions of this report appeared in Educational and Psychological Measurement, Vol. 41, No. 4 (Winter 1981).

The College Board is a nonprofit membership organization that provides tests and other educational services for students, schools, and colleges. The membership is composed of more than 2,500 colleges, schools, school systems, and education associations. Representatives of the members serve on the Board of Trustees and advisory councils and committees that consider the programs of the College Board and participate in the determination of its policies and activities.

Additional copies of this report may be obtained from College Board Publication Orders, Box 2815, Princeton, New Jersey 08541. The price is $4.

CONTENTS

ABSTRACT

Student self-selection in deciding to repeat a test was examined by contrasting the test performance of students taking the College Board's Scholastic Aptitude Test (SAT) as juniors and again as seniors with the test performance of students taking the SAT only once as juniors. Estimates of expected test performance on a common initial administration in the junior year were derived from separate equating sections and background variables. Residuals of observed minus expected test scores revealed statistically significant differences between students who took a single administration of the SAT as juniors and students who took the same initial administration but also repeated the test as seniors; the initial observed scores of students later repeating the test were consistently lower than their expected scores for both the verbal and mathematical sections. These results indicate that self-selection occurs when students decide to repeat a test and that score changes among these students reflect negative errors of measurement on the initial test administration as well as other factors.

STUDENT SELF-SELECTION AND TEST REPETITION

The extent of score change from one administration of a test to another administration of the same test is often taken as evidence of the effectiveness of a particular intervention or of the growth among certain individuals. Problems inherent to the use and interpretation of simple differences in assessing program impact or individual differences have received considerable attention. Cronbach and Furby (1970) and Linn and Slinde (1977) provide excellent critical discussions of difference scores and alternative approaches to measuring change. A further, special case in which test-retest score differences may misrepresent actual change arises when test candidates decide for themselves whether or not they should repeat a test. Under such circumstances it is to be expected that errors of measurement on the initial test administration would influence candidates' decisions regarding retesting.

Each year hundreds of thousands of applicants to schools and colleges elect to repeat an admissions test which they had taken earlier. High school students who have taken the Scholastic Aptitude Test (SAT) as juniors, for example, may decide to take the test again as seniors. Student self-selection then becomes a possible component of score change. If students decide to repeat a test because they perceive their initial scores as underestimates of their true abilities, usual assumptions about the distribution of errors of measurement on the initial test administration may not hold for this group. There would be a nonzero and presumably negative mean for the errors of measurement leading to observed scores lower than true scores. Conversely, students electing not to repeat a test would be those whose observed scores included a nonzero and positive mean for errors of measurement on the test.

This study contrasts the test performance of students taking the SAT as juniors and again as seniors with the test performance of students taking the SAT only once as juniors. Estimates of expected test performance on a common initial administration in the junior year were derived from separate equating sections and background variables. Administrations of the SAT regularly include a variable experimental section devoted to equating scores or pretesting items; scores on this experimental section do not enter into the reported verbal or mathematical scores. Thus, separate and independent equating sections provide a basis for determining whether errors of measurement in scores on reporting sections influence student decisions to retake a test.

METHOD

Samples of two groups of students were drawn from SAT history files: students who had taken the SAT only once and for the first time in their junior year and students who had taken the same initial test administration in their junior year and then repeated the test in their senior year. The administration of the SAT from May 1979 was the initial test common to the two groups as juniors, and the repeaters had also taken the SAT in November 1979 as seniors. Four of the 10 variable experimental sections randomly distributed in the common initial administration from May 1979 were verbal or mathematical equating sections, and only students whose records included these equating sections became part of the samples. Also, students in the repeater group were those who had first taken the SAT in May 1979 as juniors and again in November 1979 as seniors without any intervening administrations of the test.

Under the assumption that a student's decision to retake a test is independent of the error of his or her reporting sections, estimates of expected test performance based on equating sections and background variables for students with a single test administration should also fit the test performance of students with a subsequent, repeat test administration. The samples of students with SAT results as juniors only and students with SAT results as both juniors and seniors were split according to whether the equating section

on their initial test administration had been either a verbal or a mathematical section. Estimates of expected verbal scores from reporting sections were based on a least-squares multiple regression of observed verbal scores on verbal equating sections and background variables for students with a single test administration. The verbal equating score was expressed as a standard score, based on the particular section's mean and standard deviation, since raw or formula scores would differ from one verbal equating section to another. Background variables were taken from the Student Descriptive Questionnaire (SDQ) completed by students when registering for the SAT given in May 1979. The variables included: high school rank; years of English study; latest English grade; years of mathematics study; latest mathematics grade; educational degree aspirations; father's level of education; mother's level of education; and public/nonpublic high school. The same procedure was followed for observed mathematical scores with students who had taken a mathematical equating section.

The coefficients for each term in the above regressions, one set of coefficients for expected verbal scores and another for expected mathematical scores, were established and validated with students who had taken a single administration of the SAT in May 1979 as juniors. Roughly one-third of such students with a verbal equating section served as the sample for establishing the regression coefficients, and the other two-thirds of such students with a verbal equating section served as cross-validation samples. Because the scores of students with complete SDQ responses differ from the scores of students with incomplete SDQ responses, a maximum likelihood algorithm (Dempster, Laird, and Rubin, 1977) was used in establishing regression coefficients with incomplete data for background variables. Students who had taken a single administration of the SAT in May 1979 as juniors and had a mathematical equating section were also split into thirds for establishing and validating another set of regression coefficients for expected mathematical scores. The distribution of residuals for observed scores minus expected scores should be equivalent in the regression and cross-validation samples of students who had taken a single test administration.

Estimates of expected scores on the same initial test administration, the SAT given in May 1979, for students later repeating the test were based on these sets of regression coefficients. The group of students with a repeat test administration was split according to equating section, verbal or mathematical, and then divided again into thirds in order to check on the distribution of residuals within the group. Finally, the mean residuals between observed and expected scores on their initial test administration were compared for students with a single administration and students with a repeat test administration.

RESULTS AND DISCUSSION

A total of 253,354 test candidates took the SAT in May 1979. Most of these examinees (88 percent) were juniors in high school, and roughly one-third (32 percent) were juniors who also took the SAT in November 1979 as seniors. Approximately 32,000 examinees were juniors who took the SAT for the first and only time in May 1979 and also had a verbal or a mathematical equating section. A comparable number of examinees with a verbal or a mathematical equating section were juniors also taking the SAT for the first time but who later repeated the test in November 1979 as seniors. Table 1 shows the means and standard deviations of the test performance on reporting sections and equating sections for these groups. These descriptive statistics and all other results presented here refer to the initial SAT in May 1979 taken both by students with a single test administration as juniors and by students with the same test administration as juniors as well as a later repeat test administration as seniors. Students taking the SAT only once as juniors had slightly higher and more dispersed scores on both reporting and equating sections than did repeaters. There were also some slight differences in the descriptive profiles of the two groups: somewhat higher percentages of those students who subsequently repeated the test come from college preparatory programs, had taken three or more years of mathematics, and planned to attain at least a bachelor's degree (see Appendix A).

TABLE 1. Means and Standard Deviations of Test Performance

Group	N	SAT-Verbal Mean	SAT-Verbal sd	SAT-Mathematical Mean	SAT-Mathematical sd
Total examinees (May 1979)	253,354	432	107	478	113
Junior examinees (May 1979)	223,394	439	105	486	111
Junior repeaters (May-November 1979)	81,959	437	97	483	104
Juniors with single test administration	31,912	439.74	113.28	484.20	118.70
Verbal equating Section A	8,010	16.71	8.39		
Verbal equating Section B	8,112	15.78	8.37		
Mathematical equating Section C	7,877			10.22	6.28
Mathematical equating Section D	7,906			9.82	5.50
Juniors with repeat test administration	31,971	435.13	97.87	479.13	104.52
Verbal equating Section A	8,158	16.65	7.46		
Verbal equating Section B	8,017	15.36	7.60		
Mathematical equating Section C	8,017			10.19	5.56
Mathematical equating Section D	7,777			9.66	5.03

Correlations of equating sections and reporting sections appear in Table 2. The high correlation of the verbal equating sections, Sections A and B, with observed verbal scores and of mathematical equating sections, Sections C and D, with observed mathematical scores suggests that equating sections can provide good estimates of expected scores. Indeed, the multiple correlations resulting from a regression of observed scores on equating scores and background variables, R = .89 for SAT-Verbal and R = .88 for SAT-Mathematical (see Appendix B), barely surpass the respective simple correlations among students with a single test administration. The lower pattern of intercorrelations found in Table 2 among students with a repeat test administration compared to students with a single test administration is consistent with the somewhat lower test reliabilities for the former group (i.e., alpha reliability estimates of .91 and .93 for verbal scores and .90 and .92 for mathematical scores for the two respective groups). The standard error of measurement for verbal scores was 30 points on the 200-800 SAT scale for both groups and for mathematical scores 33 points for both groups.

TABLE 2. Correlations of Equating Sections and Reporting Sections

Group	Reporting Section	SAT-M	Equating Section A	B	C	D
Single test administration	SAT-V	.73	.88	.88	.68	.67
	SAT-M		.68	.68	.87	.86
Repeat test administration	SAT-V	.64	.85	.85	.60	.58
	SAT-M		.61	.61	.84	.83

TABLE 3. Means and Standard Deviations of Residuals from Predicted Performance

Group	SAT-Verbal N	SAT-Verbal Mean	SAT-Verbal sd	SAT-Mathematical N	SAT-Mathematical Mean	SAT-Mathematical sd
Single test administration						
Regression sample	4,497	0.18	51.65	4,374	1.29	56.63
Cross-validation sample	4,473	1.25	50.69	4,481	-0.58	55.81
Cross-validation sample	4,385	-0.86	50.49	4,332	1.07	54.77
Total	13,355	0.20	50.96	13,187	0.58	55.75
Repeat test administration						
Comparison sample	4,143	-5.62	49.25	4,129	-10.01	54.67
Comparison sample	4,186	-4.88	49.64	3,980	-8.54	55.69
Comparison sample	4,109	-4.02	49.63	4,126	-8.99	53.31
Total	12,438	-4.84	49.51	12,235	-9.19	54.55

Regression estimates of expected scores were based on the relationship of observed scores to equating scores and background variables among students who had taken the SAT only once as juniors. The coefficients for independent variables and the constant term established for calculating these regression estimates are given in Appendix B. Table 3 presents a summary of the residuals reflecting the difference between observed scores and expected scores. Because regression coefficients were based on incomplete data and residuals calculated only for students with complete data, there is a nonzero mean residual in the regression samples. Within the group of students who had taken the SAT only once as juniors in May 1979 there was no significant difference in the mean residual for the regression sample and the cross-validation samples on either verbal scores, $F(2,13352) = 1.89$, $p > .15$, or mathematical scores, $F(2,13184) = 1.48$, $p > .20$. Within the group of students who had taken the SAT for the first time in May 1979 as juniors and again in November 1979 as seniors there was no significant difference in the mean residual across three independent comparison samples on either verbal scores, $F(2,12435) = 1.08$, $p > .30$, or mathematical scores, $F(2,12232) = .782$, $p > .45$. There were, however, significant differences in mean residuals between groups for both verbal scores, $t(25791) = 8.05$, $p > .001$, and mathematical scores, $t(25420) = 14.11$, $p > .001$. The observed scores of students later repeating the test were lower than the scores expected for their initial test administration based on their performance on an equating section and their background characteristics.

These results suggest that there is student self-selection in test repetition. Apparently, students electing to repeat an admissions test do so in part because they perceive their initial test scores on reporting sections as underestimates of their true abilities. Estimates of expected scores derived from equating sections and background variables tend to confirm these student perceptions. Such self-selection in test repetition would lead to a nonzero, negative sum of errors of measurement on repeaters' initial test scores which would, in turn, distort the magnitude of score changes and preclude the application of existing models for measuring change (e.g., Lord, 1963). These findings would also seem to increase the likelihood that the student self-selection posited in other contexts (e.g., Messick, 1980) is an important factor in score change.

The amount of score change on the SAT attributable to errors of measurement remains unclear. Differences in the mean residuals reported here, five points for verbal scores and 10 points for mathematical scores, reflect both positive errors among students with a single test administration and negative errors among students with a repeat test administration, and so may represent an overestimate. Yet some students undoubtedly take the SAT only once or retake the test regardless of their initial scores. Such prejudgments

would lessen the effects of measurement error on score change. It does seem clear, however, that simple score gains or losses from one administration of an admissions test to another misrepresent change by failing to take student self-selection and other factors into account.

REFERENCES

Cronbach, L.J., and Furby, L. "How We Should Measure 'Change' — or Should We?" Psychological Bulletin, 74 (1970): 68-80.

Dempster, A.P.; Laird, N.M.; and Rubin, D.B. "Maximum Likelihood from Incomplete Data via the EM Algorithm," Journal of the Royal Statistical Society, Series B, 39 (1977): 1-38.

Linn, R.L., and Slinde, J.A. "The Determination of the Significance of Change between Pre- and Post-Testing Periods," Review of Educational Research, 47 (1977): 121-150.

Lord, Frederic M. "Elementary Models for Measuring Change." In Problems in Measuring Change, Ed. Chester W. Harris, pp. 21-38. Madison, Wisc.: The University of Wisconsin Press, 1963.

Messick, Samuel J. The Effectiveness of Coaching for the SAT: Review and Reanalysis of Research from the Fifties to the FTC. Princeton, N.J.: Educational Testing Service, 1980.

APPENDIX A: Descriptive Profiles of Student Groups

	Students with Single Test Administration (N=31,912)		Students with Repeat Test Administration (N=31,971)	
	Frequency	Percent	Frequency	Percent
Secondary School:				
Public	23,773	74.50	22,642	70.82
Nonpublic	5,213	16.34	6,348	19.86
High School Program:				
Academic (college preparatory)	22,814	71.49	25,659	80.26
General	3,907	12.24	2,380	7.44
Career (business, technical)	1,843	5.78	657	2.05
Other	109	0.34	64	0.20
High School Class Size:				
Less than 100 students	2,310	7.24	2,501	7.82
100-249 students	6,747	21.14	6,456	20.19
250-499 students	8,848	27.73	9,863	30.85
500-749 students	5,295	16.59	5,005	15.65
More than 750 students	5,372	16.83	4,809	15.04
High School Rank:				
Highest tenth	6,425	20.13	6,539	20.45
Second tenth	6,081	19.06	6,601	20.65
Second fifth	7,082	22.19	7,368	23.05
Middle fifth	6,995	21.92	6,060	18.95
Fourth fifth	781	2.45	544	1.70
Lowest fifth	128	0.40	80	0.25
Years of English:				
None	51	0.16	20	0.06
One year	208	0.65	152	0.48
Two years	268	0.84	158	0.49
Three years	1,813	5.68	956	2.99
Four years	23,795	74.56	24,575	76.87
More than four years	2,743	8.60	2,992	9.36
Years of Mathematics:				
None	92	0.29	34	0.11
One year	514	1.61	145	0.45
Two years	3,022	9.47	1,260	3.94
Three years	8,002	25.08	6,172	19.30
Four years	14,326	44.89	17,869	55.89
More than four years	2,894	9.07	3,353	10.49
Most Recent English Grade:				
Excellent (90-100, A)	9,828	30.80	9,931	31.06
Good (80-89, B)	13,420	42.05	14,435	45.15
Fair (70-79, C)	4,874	15.27	4,021	12.57
Passing (60-69, D)	561	1.76	302	0.94
Failing (below 60, F)	61	0.19	26	0.08

(continued)

APPENDIX A: Descriptive Profiles of Student Groups (continued)

	Students with Single Test Administration (N=31,912)		Students with Repeat Test Administration (N=31,971)	
	Frequency	Percent	Frequency	Percent
Most Recent Mathematics Grade:				
Excellent (90-100, A)	8,152	25.55	8,550	26.74
Good (80-89, B)	11,021	34.54	12,019	37.59
Fair (70-79, C)	7,508	23.53	6,624	20.72
Passing (60-69, D)	1,777	5.57	1,308	4.09
Failing (below 60, F)	235	0.74	141	0.44
Part-time Employment:				
None	12,405	38.87	12,260	38.35
Less than 6 hours per week	2,674	8.38	2,926	9.15
6-10 hours per week	2,976	9.33	3,434	10.74
11-15 hours per week	3,398	10.65	3,707	11.59
16-20 hours per week	3,943	12.36	3,772	11.80
21-25 hours per week	2,060	6.46	1,707	5.34
26-30 hours per week	830	2.60	614	1.92
More than 30 hours per week	396	1.24	251	0.79
Educational Aspirations:				
Two-year specialized training program	1,320	4.14	479	1.50
Two year associate's degree	788	2.47	361	1.13
Bachelor's degree	8,584	26.90	9,326	29.17
Master's degree	6,243	19.56	7,264	22.72
Professional degree	4,395	13.77	5,211	16.30
United States Citizenship:				
Yes	28,325	88.76	28,392	88.81
No	629	1.97	552	1.73
Armed Forces Veteran:				
Yes	153	0.48	154	0.48
No	28,637	89.74	28,597	89.45
Ethnic Group/National Origin:				
American Indian, Alaskan native	86	0.27	69	0.22
Black, Afro-American	1,052	3.30	1 020	3.19
Mexican-American, Chicano	142	0.44	78	0.24
Oriental, Asian-American	370	1.16	505	1.58
Puerto Rican	148	0.46	160	0.50
White, Caucasian	26,029	81.56	25,994	81.30
English as First Language:				
Yes	27,786	87.07	27,829	87.04
No	737	2.31	677	2.12

APPENDIX A: Descriptive Profiles of Student Groups (continued)

	Students with Single Test Administration (N=31,912)		Students with Repeat Test Administration (N=31,971)	
	Frequency	Percent	Frequency	Percent
Father's (Male Guardian's) Level of Education:				
Grade school	733	2.30	707	2.21
Some high school	2,124	6.66	1,832	5.73
High school diploma	6,589	20.65	5,635	17.63
Business or trade school	1,798	5.63	1,798	5.62
Some college	4,821	15.11	4,513	14.12
Bachelor's degree	4,819	15.10	5,538	17.32
Some graduate or professional school	1,397	4.38	1,643	5.14
Graduate or professional degree	6,023	18.87	6,614	20.69
Mother's (Female Guardian's) Level of Education:				
Grade school	507	1.59	497	1.55
Some high school	1,888	5.92	1,556	4.87
High school diploma	10,468	32.80	9,987	31.24
Business or trade school	2,272	7.12	2,442	7.64
Some college	5,347	16.76	5,195	16.25
Bachelor's degree	3,673	11.51	4,103	12.83
Some graduate or professional school	1,556	4.88	1,791	5.60
Graduate or professional degree	2,582	8.09	2,681	8.39
Parents' Annual Income:				
Below $3,000	184	0.58	146	0.46
$3,000-$5,999	443	1.39	416	1.30
$6,000-$8,999	524	1.64	392	1.23
$9,000-$11,999	551	1.73	506	1.58
$12,000-$14,999	856	2.68	627	1.96
$15,000-$17,999	815	2.55	680	2.13
$18,000-$20,999	1,056	3.31	881	2.76
$21,000-$23,999	895	2.80	831	2.60
$24,000-$26,999	1,326	4.16	1,143	3.58
$27,000-$29,999	1,082	3.39	985	3.08
$30,000-$34,999	1,696	5.31	1,582	4.95
$35,000-$39,999	2,160	6.77	1,884	5.89
$40,000-$44,999	1,467	4.60	1,434	4.49
$45,000-$49,999	1,793	5.62	1,763	5.51
$50,000 and over	1,125	3.53	1,127	3.53

APPENDIX B: Regression Coefficients for Estimates of Expected Scores

Independent Variables	SAT-V (N=5,602) Regression Coefficients β	B	se_B	SAT-M (N=5,436) Regression Coefficients β	B	se_B
Equating section score	0.789	85.16	0.81	0.735	83.20	0.98
High school rank	-0.075	-7.04	0.82	-0.082	-8.08	0.92
Years of English study	0.023	4.93	1.35	-0.006	-1.37	1.52
Latest English grade	-0.053	-3.60	0.53	-0.020	-1.45	0.58
Years of mathematics study	0.032	4.13	0.87	0.069	9.31	1.02
Latest mathematics grade	-0.016	-0.90	0.44	-0.058	-3.40	0.50
Educational aspirations	0.015	1.21	0.50	0.032	2.71	0.56
Father's level of education	0.029	1.60	0.39	0.040	2.25	0.44
Mother's level of education	0.023	1.42	0.44	0.021	1.35	0.50
Public/nonpublic high school	0.003	1.00	1.80	-0.011	-3.27	1.99
Constant		418.57			472.60	
Multiple correlation		0.891			0.876	
Standard error of estimate		51.538			56.733	