DOCUMENT RESUME

ED 134 610 TM 006 016

AUTHOR Breland, Hunter M.
TITLE Grade Inflation and Declining SAT Scores: A Research Viewpoint.
INSTITUTION Educational Testing Service, Princeton, N.J.
PUB DATE [Sep 76]
NOTE 21p.; Paper presented at the Annual Meeting of the American Psychological Association (Washington, D.C., September 1976)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS *College Entrance Examinations; *Educational Trends; *Grades (Scholastic); *Higher Education; Low Achievement Factors; Predictive Validity; *Scores; Secondary Education
IDENTIFIERS ACT Assessment Program; *Grade Inflation; Scholastic Aptitude Test; *Test Score Decline

ABSTRACT

Evidence indicates that college grades have increased on the average and that the college-bound population has decreased in traditional kinds of academic skills, based on observed declines on the Scholastic Aptitude Test (SAT) and the American College Test (ACT). But the relationship between these traditional skills and grades appears not to have changed. It is speculated that, despite the increase in average grades, the reliability of the grades awarded by college professors has not diminished. If either grades or test scores had decreased in reliability, then the correlations between the two would probably have decreased. The test score reliability is carefully maintained. Whatever the case, the impact of grade inflation and score decline can probably be anticipated. Already there is evidence that grades are beginning to deflate. And with the wide-spread public concern about test score declines, it may be expected that these will begin to level off and even to rise in the next few years.

GRADE INFLATION AND DECLINING SAT SCORES:

A RESEARCH VIEWPOINT

Hunter M. Breland

Educational Testing Service

Princeton, New Jersey

Paper prepared for presentation at the Annual Meeting of the American Psychological Association, Washington, D. C., September, 1976.

Grade Inflation and Declining SAT Scores:

A Research Viewpoint

Hunter M. Breland
Educational Testing Service

According to a recent study by the Office of Institutional Research at the University of California at Berkeley, the number of A's awarded at a group of colleges studied more than doubled between the early 1960's and the early 1970's--while the number of C's decreased by almost 50 percent. In this study, the percentage of A's increased from 16 percent of all grades in the early 1960's to 34 percent in the early 1970's, the percentage of C's decreased from 37 percent to 21 percent, and the average grade point index for the responding institutions increased from 2.49 in 1963 to 2.94 in 1974 (Chronicle of Higher Education, March 22, 1976).

Arvo Juola, of Michigan State University, conducted a study at 197 institutions and found that, since 1960, grade point averages had increased abou one-half of a letter grade, on the average. Benno Fricke reported a study at the University of Michigan showing that the freshman GPA's at that institution for the academic year 1974-75 were the highest ever earned--despite the fact that this same freshman class was the weakest in more than two decades in terms of its performance on standardized tests (Chronicle of Higher Education, May 19, 1975).

Eighty-two percent of the 1974 graduating class at Harvard graduated cum laude or better. Recently 81 percent of all grades at Vassar were A's or B's, the average grade at Stanford was A-minus, and more than half the University of Virginia student body made the Dean's List. Over a ten-year period, the percentage of A students at the University of North Carolina doubled and the average grade at the University of Wisconsin rose from C-plus to B-plus (Newsweek, February 9, 1976).

Phi Beta Kappa reports, that to keep their membership selective, many of the 214 chapters are raising the minimum requirements. One chapter (at the University of Indiana) refused to admit new students in the spring of 1976 until selection standards are stiffened (Chronicle of Higher Education, April 26, 1976).

Dickinson College has for a number of years published a listing of students who made grade point averages of 3.50 or better. The publication of this listing has recently been temporarily discontinued because its length has become an embarrassment (Higher Education Daily, January 23, 1975).

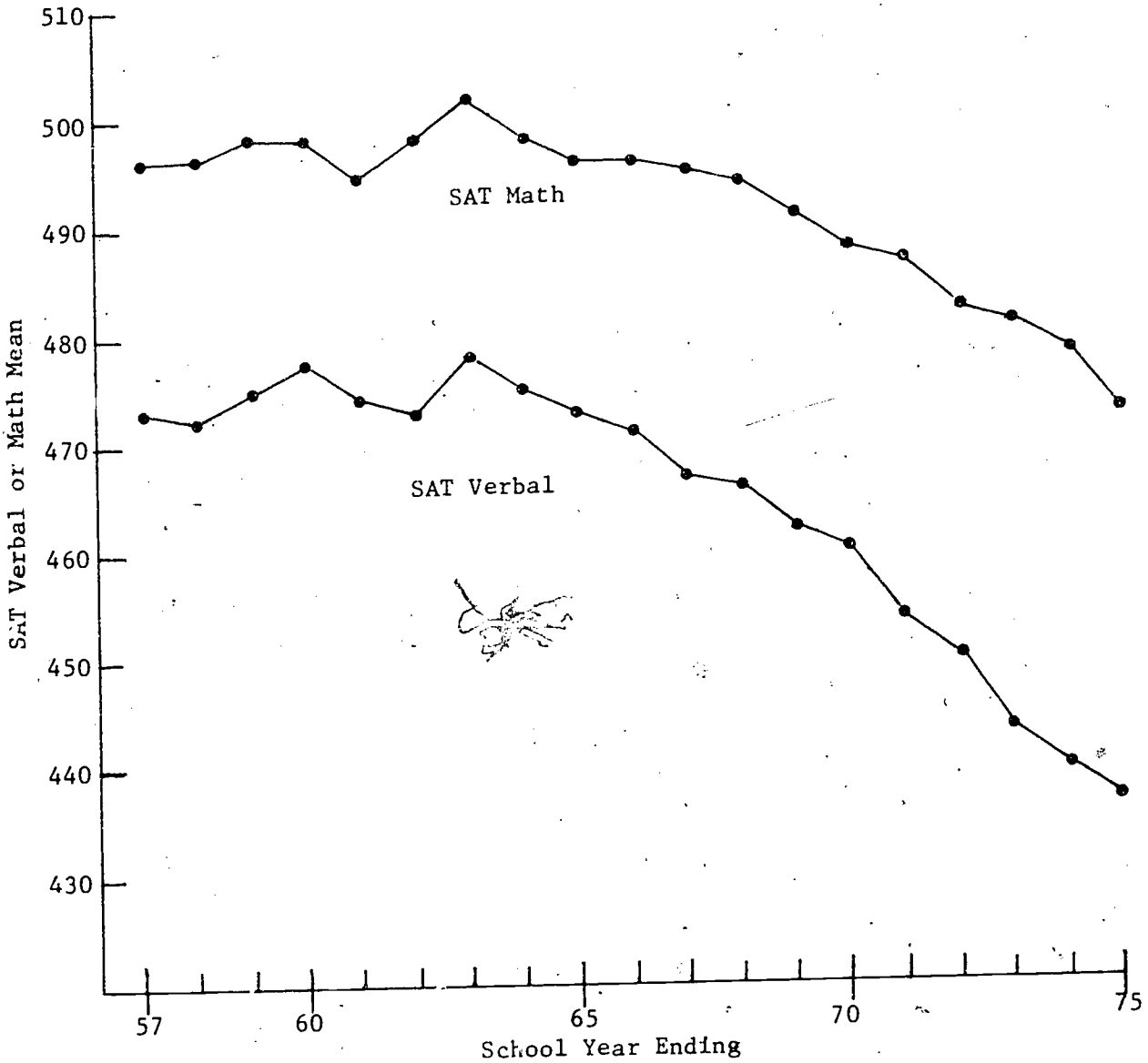
The SAT Score Decline

In sharp contrast to this apparent inflation of grades, average scores on the College Board Scholastic Aptitude Test (SAT) have declined each year for twelve successive years since 1963. Verbal score averages have declined a total of more than 30 points and a similar, but less dramatic, decline has occurred in the mean SAT mathematical score. The nature of these declines is depicted in Figure 1. Table 1 presents the same data in tabular form. During the past two or three years numerous explanations have been offered for these declines. Almost a year ago, the College Board named a panel of experts to review the score decline and to determine which explanations seem most promising.

All of the explanations relate to some kind of change that has occurred concurrently with the score declines, and these explanations fall into four general areas of classification: the test, the population taking the test, the schools, and societal factors. Explanations suggesting that the problem resides within the test, although less common than other explanations, usually point to the possibility that the test may have become more difficult. It is also sometimes suggested that tests like the SAT are simply irrelevant in a modern world where written materials are rapidly being replaced with other communication media.

Figure 1

SAT Verbal and Math Means for the Period 1957 - 1975*



*These means represent all scores reported. Thus, persons having more than one score are counted more than once.

Table 1. SAT Verbal and Mathematical Means

Testing Year	SAT Verbal and Mathematical Means	
	V	M
1956-57	473	496
1957-58	472	496
1958-59	475	498
1959-60	477	498
1960-61	474	495
1961-62	473	498
1962-63	478	502
1963-64	475	498
1964-65	473	496
1965-66	471	496
1966-67	467	495
1967-68	466	494
1968-69	462	491
1969-70	460	488
1970-71	454	487
1971-72	450	482
1972-73	443	481
1973-74	440	478
1974-75	437	473

[a]These are the values presented in Figure 1. These means represent all scores reported. Students represented in these means are aggregated without regard to their grade level, and students are counted as many times as they have taken the SAT.

Others argue that, because of emerging egalitarian philosophies--coupled with financial difficulties in postsecondary institutions--many colleges are no longer as selective as they used to be. As a result, students may have been admitted to colleges in the past decade who would have been rejected in previous years. The opportunity to go to college has been extended to increasingly large segments of the population. In fact, students today are aggresively recruited for college. A symptom of this trend in college admissions, it is alleged, is a trend toward increasingly easier college textbooks, since it is claimed that today's college students cannot read as well as did students of earlier years. Of course, the population of students taking the SAT may not be representative of the college-bound population, and changes in the SAT candidate population over the period of the declines may have been different from changes occurring in the total college-bound population.

The most popular set of explanations shifts the burden from the colleges to the secondary schools. In particular, there is speculation about the quality of secondary education in America. In their attempts to be progressive, it is said, schools have become too permissive, and the result has been that traditional academic skills are no longer learned. Those who espouse this explanation often note the increasing constraints on the teaching profession, the encroachment of the courts, and reduced expenditures for education.

A final set of explanations focuses on a broad range of societal factors. Advocates of these explanations note trends among adolescents toward drug use, alcoholism, television watching, and other similar factors which would appear to have detrimental effects on the development of academic aptitudes. Additionally, some would relate the decreased competition for college admission and the use of non-academic criteria in awarding scholarships to a reduction in the motivation to perform well on admissions tests.

Although it has not normally been the focus of explanations, it is appropriate to ask whether some change in the instrument itself might have contributed to the score decline. If the SAT items contain material no longer normally learned, either at home or in school, then it would be expected that recent cohorts would perform poorly. Or, it could be, in an advanced technological society dominated by television and other non-written forms of communication, skills such as those examined in the SAT are no longer necessary or relevant.

Since the SAT is changed each year as different forms are generated and since it has changed over the years to make it more responsive to the needs of the educational community and to improve individual questions where possible, there has always been the concern that the test could become either slightly more difficult or slightly easier as such changes are introduced. Because of this concern, a complex and sophisticated scaling procedure has been developed over the years to maintain a scale of constant meaning.

A clarification of the difficulty issue is provided by the performance of samples on test sections used for equating earlier test forms to new forms--that is, sections that are identical across test administrations and testing years. During the past decade, new samples of the SAT population have consistently scored lower on the same material than old samples. Data from one study, is presented in Table 2.

In the case of the first verbal comparison of Table 2, a reprinted section of a verbal test given in 1963 was readministered in 1973--and in the same month, December. Out of the 40 items in this reprinted section, the 1963 sample averaged in excess of 18 (raw score points[1]); the

---

[1]Raw score = Number correct minus 1/4 of the number incorrect.

1973 sample averaged less than 14. For the shorter period from 1966 to 1973, the verbal decline is slighly less than that from 1963 to 1973, as would be expected. The same pattern is evident in the comparisons of the mathematical sections in Table 2. In 1963, the sample obtained a raw score average of more than 10 on the 25 mathematical items. In 1973, when the same 25-item section was reprinted and administered, the sample obtained a raw score average of less than 8. And the mathematical decrease from 1966 to 1973 is slightly less than the mathematical decrease for the longer period from 1963 to 1973.

Table 2. Comparison of Identical SAT Test Sections for Two Time Intervals[1]

Test Section	Number of Items	When Administered	Sample Size	Mean Raw Score	Standard Deviation
Verbal	40	Dec. 1963	1,500	18.43	8.13
Verbal	40	Dec. 1973	5,000	13.99	7.92
Verbal	40	Dec. 1966	5,000	17.23	8.19
Verbal	40	Dec. 1973	4,915	13.99	8.20
Math	25	Dec. 1963	1,500	10.41	5.54
Math	25	Dec. 1973	5,000	7.90	5.38
Math	25	Dec. 1966	5,000	10.94	5.70
Math	25	Dec. 1973	4,921	8.91	5.67

[1]From Modu and Stern (1975)

The ACT Score Decline

To a degree, the SAT population represents the college-bound population, but it must be recognized that not all persons seeking college admission take the SAT. Many institutions require the tests administered by the American College Testing Program (the ACT), other institutions allow an option of the SAT or the ACT, and some institutions require neither. Most college-bound students, however, take either the SAT or the ACT, and some take both. In examining the college-bound population, therefore, the experiences of the ACT are of special interest, since if

both testing programs have had similar experiences, then those experiences would be representative of the college-bound population. On the other hand, if these two national testing programs have had different experiences, then the causes would seem to reside in the population tested in one program or the other and not in the total college-bound population.

Consider the decline in mean scores on the test administered by the American College Testing (ACT) program shown in Table 3. Between 1965 and 1974, the mean ACT English score declined from 18.7 to 17.6 (1.1 ACT scale points), the mean ACT Mathematics score from 19.6 to 18.1 (1.5 ACT scale points), and the mean ACT Social Studies score from 20.6 to 17.9 (2.7 ACT scale points). The fourth part of the ACT test, Natural Science Reading, did not show a significant change in mean score. The four ACT subscores are combined to form what is known as the ACT Composite score. For 1975, the ACT Composite score mean was reported to be 18.6--a decline of 1.3 ACT scale points since 1965, when the ACT Composite score mean was 19.9. The SAT standard deviation is about 110 and the ACT Composite standard deviation is about 5.5. Therefore, the SAT decline between 1965 and 1975 is about 29.5/110 = .27 of a standard deviation and the ACT mean Composite score decline is about 1.3/5.5 = .27 of a standard deviation.[2] The overall declines in SAT and ACT means would appear to be of the same order of magnitude. The SAT Verbal score means and the ACT Social Science Reading score declined more than did the overall means.

Since the majority of the students in the college-bound population take either the SAT, the ACT, or both, these declines are, essentially, representative of the college-bound population.

---

[2]The SAT decline of 29.5 was derived as the average of the Verbal and Mathematical declines.

Table 3. Means and Standard Deviations for College-Bound Students Taking the ACT Test in Successive Years[a]

Year	English		Mathematics		Social Studies		Natural Science		Composite	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
1964-65	18.7	5.1	19.6	6.6	20.0	6.4	20.4	6.1	19.9	5.2
1965-66	19.1	5.1	19.5	6.7	20.5	6.4	20.5	6.1	20.0	5.2
1966-67	18.5	5.3	18.7	7.0	19.6	6.6	20.1	6.3	19.4	5.4
1967-68	18.1	5.3	18.3	7.3	19.4	6.7	19.8	6.5	19.0	5.5
1968-69	18.4	5.2	19.2	6.9	19.4	6.7	20.0	6.4	19.4	5.3
1969-70	18.1	5.3	19.5	6.7	19.3	6.8	20.5	6.1	19.5	5.3
1970-71	17.7	5.6	18.7	7.2	18.3	7.2	20.2	6.4	18.9	5.6
1971-72	17.6	5.6	18.6	7.3	18.4	7.3	20.3	6.5	18.8	5.7
1972-73	17.8	5.4	18.8	7.2	18.1	7.5	20.5	6.5	18.9	5.8
1973-74	17.6	5.4	18.1	7.5	17.9	7.6	20.6	6.5	18.7	5.8
1974-75									18.6[b]	

[a]From Ferguson and Maxey (1975).

[b]From ACTivity, October, 1975.

Effects on SAT Validity

The inflation of grades has led some to speculate that the predictive validity of the SAT (for predicting freshman GPA) may have declined. There is no evidence to support such a conclusion, however. In fact, so long as the range of grades given is not severly restricted, or the reliability of grades has not changed, there is no reason to suspect that predictive validities would have changed. In fact, if the rank order of students on SAT scores were the same as the rank order of students on grades, the correlation would be the same even if the means on grades increased and the means of SAT decreased. Grade inflation has, essentially, the effect of adding a constant to student GPA and the adding of a constant to either of two variables being correlated has no effect on the correlation coefficient.

Tables 4 and 5 provide evidence that: (a) grades have inflated, but (b) that SAT predictive validities have not changed. These tables show changes in the correlation between SAT scores and GPA over a number of years for 27 colleges. In 15 of the 27 colleges, higher predictive validities were obtained for their most recent classes than were obtained for earlier classes. Table 5, where the colleges are ranked in descending order of grade inflation, shows no particular pattern of correlational changes with grade inflation changes. Table 6, adapted from the 1975-76 Guide for High Schools and Colleges (CEEB, 1975), shows also that there are no appreciable differences in predictive validities across various college curriculums. Table 7, a revision of Table 6 based on more recent data, shows no substantial differences.

Numerous studies published in professional journals have also demonstrated the predictive validity of the SAT during the period of the score decline (e.g., Cleary, 1968; Temp, 1971; Fincher, 1974). One is led to conclude, from both these published studies as well as yet unpublished

Table 4 Comparison of Grade Inflation and Validity Changes in 27 Institutions

(Adapted from Ford, 1976)

College	Mean Freshman GPA		Correlations		Differences Between		Student Year of Entrance	
	Most Recent	Earliest	Most Recent	Earliest	Means	Correlations	Most Recent	Earliest
A	2.67	2.05	.44	.35	.62	.09	72	67
B	2.19	2.00	.54	.49	.19	.05	71	64
C	2.75	2.35	.43	.39	.40	.04	72	65
D	2.72	2.54	.56	.44	.18	.12	72	65
E	2.49	2.26	.41	.48	.23	-.07	72	69
F	3.12	2.72	.44	.24	.40	.20	71	67
G	2.64	2.48	.34	.35	.16	-.01	72	66
H	2.58	2.05	.22	.14	.53	.08	72	67
I	2.46	2.24	.42	.46	.42	-.04	72	64
J	2.65	2.28	.30	.15	.37	.15	72	66
K	2.43	2.24	.25	.46	.19	-.21	72	66
L	2.47	2.05	.37	.44	.42	-.07	72	67
M	2.32	2.23	.40	.38	.09	.02	71	64
N	2.37	1.80	.28	.40	.57	-.12	71	65
O	2.12	1.72	.41	.51	.40	-.10	72	65
P	2.50	2.20	.49	.34	.30	.15	72	65
Q	2.47	2.04	.51	.44	.43	.07	71	66
R	2.24	2.20	.41	.47	.04	-.06	72	68
S	2.28	2.21	.34	.38	.07	-.04	71	65
T	2.43	2.39	.35	.34	.04	.01	72	67
U	2.29	2.11	.42	.56	.08	-.14	71	66
V	2.61	2.19	.42	.41	.42	.01	72	63
W	2.17	1.69	.29	.43	.48	-.14	71	66
X	2.25	1.87	.34	.37	.38	-.03	72	64
Y	2.09	1.93	.63	.52	.16	.11	72	67
Z	2.88	2.40	.59	.48	.48	.11	71	64
AA	2.64	2.40	.40	.32		.08	72	65

Note: Correlations are those observed between GPA and composite SAT scores $(\frac{2V + M}{3})$.

Table 5 Comparison of Grade Inflation and Validity Changes in Rank-order of Mean GPA Differences

(Adapted from Ford, 1976)

College	Differences In Means	Differences In Correlations	Years
A	.62	.09	5
N	.57	-.12	6
H	.53	.08	5
W	.48	-.14	5
Z	.48	.11	7
Q	.43	.07	5
I	.42	-.04	8
L	.42	-.07	5
V	.42	.01	9
C	.40	.04	7
F	.40	.20	4
O	.40	-.10	7
X	.38	-.03	8
J	.37	.15	6
P	.30	.15	7
AA	.24	.08	7
E	.23	-.07	3
B	.19	.05	7
K	.19	-.21	6
D	.18	.12	7
G	.16	-.01	6
Y	.16	.11	5
M	.09	.02	7
U	.08	-.14	5
S	.07	-.04	6
R	.04	-.60	4
T	.04	.01	5

Note: Correlations are those obtained between GPA and composite SAT scores $(\frac{2V + M}{3})$.

Table 6. Predictive Validity Coefficients for SAT Scores and High School Records of Students in Various College Curriculums

Classification of Groups	Number of Groups	SAT-Verbal			SAT-Mathematical			High School Record			Combined SAT Scores and School Record		
		90%ile	Mdn.	10%ile	90%ile	Mdn.	10%ile	90%ile	Mdn.	10%ile	90%ile	Mdn.	10%ile
Liberal Arts Men	116	.48	.33	.17	.47	.30	.16	.62	.47	.31	.68	.55	.40
Liberal Arts Women	143	.55	.41	.23	.52	.36	.21	.68	.54	.35	.74	.62	.43
Liberal Arts Men and Women	51	.54	.39	.26	.48	.33	.20	.67	.55	.33	.73	.62	.46
Engineering	23	.37	.23	.11	.41	.33	.19	.53	.42	.26	.60	.51	.34
Science	10	.45	.34	.24	.44	.32	.24	.60	.54	.37	.65	.58	.44
Business	17	.36	.30	.13	.37	.29	.21	.54	.39	.28	.62	.48	.40
Education	13	.52	.35	.31	.42	.37	.26	.64	.50	.38	.74	.58	.49

Source: 1975-76 ATP Guide for High Schools and Colleges (CEEB, 1975).

Table 7 Predictive Validity Coefficients for SAT Scores and High School Records of Students in Various College Curriculums

Classification of Groups	Number of Groups	SAT-Verbal			SAT-Mathematical			High School Record			Combined SAT Scores and School Record		
		90%ile	Mdn.	10%ile	90%ile	Mdn.	10%ile	90%ile	Mdn.	10%ile	90%ile	Mdn.	10%ile
Liberal Arts Men	101	.45	.35	.20	.46	.37	.22	.56	.45	.28	.62	.52	.38
Liberal Arts Women	116	.55	.41	.30	.55	.41	.25	.63	.49	.31	.70	.57	.44
Liberal Arts Men and Women	145	.56	.42	.27	.55	.40	.25	.65	.63	.37	.71	.61	.45
Engineering	16	.47	.31	.13	.52	.38	.30	.57	.48	.36	.65	.56	.48
Science	18	.55	.31	.22	.59	.41	.28	.64	.47	.40	.69	.56	.48
Business	22	.49	.34	.23	.52	.35	.22	.64	.52	.35	.67	.60	.42
Education	14	.52	.40	.28	.53	.40	.31	.66	.53	.36	.70	.63	.43

Source: 1976-77 Guide for High Schools and Colleges (CEEB, 1976, in press).

analyses, that the grade inflation has had no observable effect on the validity of the SAT. Analyses of this type, of course, are conducted on a regular basis as an integral part of College Board programs.

Conclusions

What, then, does all of this signify? The evidence shows that college grades have indeed increased, on the average. The evidence suggests that the college-bound population has decreased in traditional kinds of academic skills. But the relationship between these traditional skills and grades appears not to have changed. One is led to speculate that, despite the increase in average grades, the reliability of the grades awarded by college professors has not diminished. If either grades or test scores had decreased in reliability, then the correlations between the two would probably have decreased. The test score reliability is carefully maintained.

Whatever the case, the impact of grade inflation and score decline can probably be anticipated. Already there is evidence that grades are beginning to "deflate." One study, at the University of Minnesota, indicates a decrease in the percentage of high grades awarded (Chronicle of Higher Education, December 22, 1975). And with the wide-spread public concern about test score declines, one may expect that these will begin to level off and even to rise in the next few years.

References

Cleary, T.A. Test bias: Prediction of grades of Negro and White students in integrated colleges. Journal of Educational Measurement, 1968, 5, 115-124.

College Entrance Examination Board (CEEB). Guide for high schools and colleges. Admissions Testing Program of CEEB, 1975-76.

College Entrance Examination Board (CEEB). Guide for high schools and colleges. Admissions Testing Program of CEEB, 1976-77 (in press).

Ferguson, R.L., and Maxey, E.J. Trends in the academic performance of high school and college students. (Unpublished manuscript), Iowa City, Iowa: The American College Testing Program, 1975.

Fincher, C. Is the SAT worth its salt? Review of Educational Research, Summer 1974, 44, (3), 293-305.

Ford, S.F. Grade inflation and validity of the SAT. (Unpublished manuscript), Educational Testing Service, Princeton, New Jersey, April 1976.

Modu, C.C., and Stern, J. The stability of the SAT score scale. (ETS Research Bulletin, RB #75), Educational Testing Service, Princeton, New Jersey, April 1976.

Temp, G. Test bias: Validity of the SAT for Blacks and Whites in thirteen integrated institutions. Journal of Educational Measurement, 1971, 8, 245-251.