DOCUMENT RESUME

ED 258 984 TM 850 324

AUTHOR	Powers, Donald E; And Others
TITLE	Test Score Changes on the GRE General (Aptitude) Test.
INSTITUTION	Educational Testing Service, Princeton, N.J.
REPORT NO	ETS-RR-85-4
PUB DATE	Jan 85
NOTE	42p.
PUB TYPE	Reports - Research/Technical (143)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	Aptitude Tests; College Admission; *College Entrance Examinations; Higher Education; Scores; Standardized Tests; *Student Attitudes; *Test Coaching; Test Use
IDENTIFIERS	Graduate Record Examinations; Score Variation; *Test Repeaters

ABSTRACT

Test repetition is a significant phenomenon in standardized admissions testing in terms of the numbers of examinees involved. This study was initiated to determine what additional advice might be offered to test takers contemplating repetition of the Graduate Record Examinations (GRE) and test users confronted with how to interpret multiple test scores in the admissions process. The study provided additional support for advising test takers of the desirability of test preparation. Several techniques for evaluating multiple test scores were presented including use of the highest score, the most recent score, or the average of all test scores. The average of several scores, if earned in a short period of time, may be the best technique. However, regardless of the approach adopted, it should be used consistently with all applicants. Through a survey of examinees who had repeated the GRE General (Aptitude) Test, this study documented some of the factors involved in GRE test takers' decisions to repeat the test and examined the relationship of these factors to test score changes. (DWH)

RR-85-4

RESEARCH REPORT

TEST SCORE CHANGES ON THE GRE GENERAL (APTITUDE) TEST

Donald E. Powers
Mary Jo Clark

in collaboration with Jerilee Grandy

Educational Testing Service
Princeton, New Jersey
January 1985

FINAL REPORT

Test Score Changes on the GRE General (Aptitude) Test

Donald E. Powers

Mary Jo Clark

in collaboration with Jerilee Grandy

Abstract

In terms of the numbers of examinees involved, test repetition is a significant phenomenon in standardized admission testing. Although it is readily acknowledged that those who repeat admissions tests are a self-selected group of test takers, there are continuing questions about the bases on which examinees decide to retake a test.

Through a survey of examinees who had repeated the GRE General (Aptitude) Test, this study documents some of the factors involved in GRE test takers' decisions to repeat the test and examines the relationships of these factors to test score changes. Implications are drawn for advising examinees who contemplate retaking the GRE as well as graduate admissions staff who are confronted with multiple test scores.

Test Score Changes on the GRE General (Aptitude) Test

Changes in the test scores of examinees who take admissions tests on more than one occasion are of interest both to test takers and to those who evaluate and interpret test scores. In particular, test candidates may desire information for decisions about whether or not to retake a test, and test users may seek advice on how to treat multiple test scores in the admissions process. Examinees are typically interested in the likelihood of increasing their test scores on retesting, and in the effects of any score changes on the chances of gaining admission to graduate school. Admissions officers, or others who use test scores, strive to make the best decisions possible about who should be admitted.

If gauged in terms of the numbers of test takers involved, test repetition is a significant phenomenon in graduate admission testing. For the General (Aptitude) Test of the Graduate Record Examinations (GRE) Board, for instance, about 11% of all examinees have in recent years indicated that they had taken the test at least once during a previous testing year (Wild, 1981; Goodison, 1982, 1983). Because some examinees do not acknowledge having previously taken the test and because additional numbers of test takers will eventually repeat the test at some future date, 11 percent may underestimate the actual proportion of test repeaters for the GRE General Test.

Test score changes can pose problems because of the inferences made from such changes, both by test takers and by test users. Examinees may attribute score differences, particularly gains, to intervening experiences that they judge to be relevant--e.g., programs of special test preparation

or coaching. And, those who use tests may assume that large score changes, especially those spanning significant periods of time, indicate real growth in the abilities measured by the test. To others, large score gains or losses may signal that tests are inherently unreliable. As one examinee in our sample put it, "I raised my quantitative score 140 points, but unfortunately my verbal score fell. Just shows that scores are random and can be easily manipulated." The major objective of the study reported here was to learn more about the bases on which test candidates decide to repeat a test and, if possible, to relate these factors to test score changes.

Previous Research

Research on test repetition has been conducted for several major admissions testing programs, including those sponsored by the College Board, the Graduate Record Examinations Board, the Graduate Management Admission Council, and the Law School Admission Council. For these programs investigators have sought to provide:

(a) documentation of test score performance relative to such variables as the frequency of test repetition and the length of time between test administrations (e.g., Kingston & Turner, 1984; Pitcher, 1966; Rock & Werts, 1980)

(b) evaluation of the effects on score changes of specific intervening experiences, such as test disclosure (Stricker, 1982), test practice (Levine & Angoff, 1958), and special test preparation (Leary & Wightman, 1983)

(c) estimates of the reliability and validity of initial and subsequent test scores (e.g., Boldt, 1977; Linn, 1977; Olsen & Schrader, 1959; Pitcher, 1977, and Watkins & Schrader, 1963) and

(d) explanations of test score changes in terms of such factors as self-selection, growth in abilities, and measurement error (Alderman, 1981a, 1981b; Campbell, Hilton & Pitcher, 1967; Jacobs, 1966).

Because different tests have been studied in these investigations, it is difficult to say with any certainty which of the findings are test-specific and which may apply more generally to standardized admission tests. However, one finding that seems likely to apply generally is that test repeaters are a self-selected group of test takers. This conclusion has been reached, for instance, both in large-scale statistical studies of the Scholastic Aptitude Test (SAT) and the GRE General (Aptitude) Test and in smaller-scale studies of the same tests. For example, by using separate, concurrently-administered equating forms of the SAT (for which scores are not reported), Alderman (1981a) demonstrated definite self-selection effects and negative errors of measurement in initial scores. That is, the initial test scores of repeaters were systematically lower than their estimated true scores, suggesting that self-selection is due in part to examinees' perceptions of their initial scores as underestimates of their true abilities, possibly because of disparities between scores on various sections of the test.

Rock and Werts (1980) showed that individuals who retook the GRE General (Aptitude) Test more than once were, on average, of lower ability

than those who retested only once. Campbell et al. (1967) found that substantial numbers of GRE test repeaters felt that their initial test performances were not adequate reflections of their true capabilities, and Jacobs (1966) discovered that examinees attributed SAT score changes to such factors as poor health, confidence/nervousness, and concurrent enrollment in mathematics courses.

However, although the evidence is substantial that test repeaters are self-selected, we know considerably less about the bases for their self-selection. Besides the factors that have been isolated in the research cited here, test repeaters are undoubtedly self-selected in many other ways that remain largely undocumented. The purpose of this study was to provide additional documentation of some of these factors.

Procedures

Sample Selection

A sample of 1,543 prospective test takers was chosen from those who registered for the June 1980 GRE-national administration and who identified themselves as having taken the test before; the sample included all Black registrants and a spaced sample of all White registrants who identified themselves as test repeaters. Because 432 of these test registrants had also been selected for a concurrent GRE study of the effects of special test preparation, they were deleted from the study sample, leaving a total of 1,111 candidates. Each of these remaining registrants was mailed a questionnaire that sought information on (a) perceived factors that affected test-taking performances on each occasion, (b) intervening experiences that might relate to test score changes, and (c) reasons for retaking the test.

A total of 864 of these prospective test takers took the GRE Aptitude (General) Test in June 1980 and received test scores. However, previous test scores could be located for only 580 of these examinees, even after a manual search of GRE microfilm records. The apparent discrepancy between examinee reports and test records can be explained by such factors as (a) examinees changing names between testings, which precluded the easy retrieval of previous scores (b) lack of information about the date of previous testing, which is needed to facilitate searches of microfilm records; and (c) test takers misreporting prior test taking experience, e.g., indicating they had previously taken the Aptitude (General) Test when in fact they had not or had taken only an Advanced (Subject) Test. Because of these factors, women and older test takers (who were more likely to have taken the previous test several years earlier) were over-represented in the group for whom no second test score could be found.

Questionnaires were mailed immediately after the test administration so that the test-taking experience would still be fresh in the minds of examinees. Most responses were returned prior to the mailing of test scores, which took place about a month after the test administration. Of the 1,111 test takers who were surveyed, a total of 716 returned usable questionnaires after initial nonrespondents had been recontacted. Some respondents indicated that they had not in fact previously taken the test, and were subsequently deleted from the sample. A total of 628 respondents said they had taken the test on at least one prior occasion. These respondents represented 73% of the 864 members of the sample who took the June 1980 test.

A cross tabulation of test takers and questionnaire respondents is presented in Table 1. As can be seen, a core sample of 433 test repeaters was available for the analysis of questionnaire responses in relation to changes in test scores. When only questionnaire data were analyzed, the larger total of 628 questionnaire respondents was included in order to make maximum use of our data.

Insert Table 1 about here

Results

Table 2 presents a comparison of the primary study sample with the population of GRE test takers in 1979-80, the testing year in which the data for this study were collected. As is evident, the test takers in our sample differed from the GRE population in several respects. Compared with most GRE test takers, they tended (a) to identify themselves as either White or Black, with a higher proportion of Blacks; (b) to have higher degree objectives; and (c) to have received their undergraduate degrees less recently.

The relatively large proportion of Blacks (21% vs. 7% among all test takers) resulted from intentionally oversampling Black test takers when selecting the sample, as previously described. Because, the results of this study were quite similar for Blacks and Whites, the data are not presented separately. However, we note that because the test scores of Black test takers were about one standard deviation below the average scores for White test takers, the mean "Previous" and "Recent" GRE scores reported here are lower (by nearly 30 points) than if Black test takers had not been oversampled. Even so, the sample of repeaters earned somewhat

lower initial test scores, particularly on the quantitative and analytical sections. Considering the different mixes, it appears that the sample of repeaters earned similar or slightly higher verbal and analytical retest scores when compared to all 1979-80 test takers, and only slightly lower quantitative scores. The profile of all 628 questionnaire respondents is similar to that of the primary study sample, except that there are more women, more older test takers, and no available earlier test scores in the larger, respondent sample.

Insert Table 2 about here

The major results of the study will be presented here as the answers to a series of questions pertaining to various aspects of the test repetition phenomenon.

1. How frequently do examinees retake the GRE Aptitude Test?

Of 628 questionnaire respondents, 89% indicated taking the examination twice, 8% three times, and nearly 3% more than three times. Table 3 shows the self-reported time periods between the June testing and the most recent previous test. As shown, the greatest percentage of examinees repeated the test between 6 months and 2 years or between 3 and 8 years. According to examinee reports, however, a significant number (22%) retook the test after 9 years or more, and these are the examinees whose test scores were most difficult to retrieve. Thus the sample on which most of our analyses are based is biased towards excluding older test takers who repeated the test after a significant interval of time.

Insert Table 3 about here

2. What are the patterns of GkE score changes?

Table 4 shows the patterns of test score increases and decreases for the test repeaters in the sample. By far the greatest percentage of both large and small gains and smallest percentage of decreases were for the analytical section of the test, which was revised after this research began, when research revealed that this section of the test was susceptible both to short term practice and to special test preparation. The sizable gains for the analytical section (Mean = 56.7; 29.3% gained 100 points or more) are due in large part to this susceptibility.

Insert Table 4 about her

The average change on the quantitative section was less than on any other section of the test, but score differences were also more variable for this section than for any other. Over a third (35.6%) of the test takers obtained lower quantitative scores on retesting than on the initial test, and for about one of every six test takers, quantitative scores decreased by 50 points or more. Gains in verbal scores averaged about 31 points, and although verbal score changes were less variable than those for any other section, nearly one in seven candidates gained 100 points or more and at least one in four test takers exhibited a test score decrease. Despite differences in samples, these results are generally similar to those reported by Rock and Werts (1980).

3. How do test score changes relate to intervals between tests?

As might be expected, the length of time between test administrations correlated significantly ($r = .60$) with the number of years since receipt of undergraduate degree. The correlations of months between tests with test score change were slight for each section of the test: verbal, $r = .14$; quantitative, $r = -.12$; and analytical, $r = .09$. Although slight, these correlations are significant ($p < .05$) with the relatively large samples employed here. These results are generally consistent with those of Rock and Werts (1980), who found a small positive relationship between time lapse and verbal score gain but essentially no relationship for quantitative scores. (It should be noted that Rock and Werts studied the relationship of time lapse to retest scores, controlled for initial scores for periods up to three years.) These results suggest that, as Rock and Werts (1980) noted, verbal scores may increase with the everyday acquisition of verbal knowledge, but quantitative ability/achievement is less likely to improve with the simple passage of time. This conclusion is also supported by an examination of our data grouped by time intervals between tests. This analysis revealed larger average increases in verbal scores in each time interval longer than three years, while the larger average increases in quantitative scores occurred in time intervals shorter than three years.

Because the analytical measure was not introduced until October 1977, the longest possible interval between analytical tests was less than three years. Conclusions about changes in analytical scores over time were, therefore, somewhat more difficult to draw because of the more restricted time frame for this section of the test.

The levels of test scores for persons of different ages are also often of interest, in addition to score changes that might be anticipated for various intervals of times between tests. Table 5 presents mean scores by self-reported time between tests for the 433 members of the sample who had test scores for both the June 1980 and a previous test administration. As can be seen, those electing to retake the test after three years or more had higher previous verbal and quantitative scores than did those who were repeating sooner, suggesting that the older repeaters were a more highly self-selected group. The older test takers also gained more on the verbal section than did those who were repeating within two years, but gained less on the quantitative section.

Insert Table 5 about here

The pattern of GRE scores by length of time between tests is generally consistent with recent analyses of the GRE scores of younger and older test takers (Clark, 1984; Hartle, Baratz, & Clark, 1983). These studies examined the scores of test takers grouped by age (22 or less, 23-29, 30-39 and 40 or more) and by year since undergraduate degree (9-15 years vs. 16 or more years beyond the bachelor's degree). The results of these studies indicated that the average verbal scores were about the same for all test takers (who were not necessarily repeating the test) regardless of age or recency of undergraduate study, while average quantitative scores were progressively lower with increasing age and time since the baccalaureate.

4. What reasons do examinees give for retaking the test?

When asked why they had retaken the test, a majority of repeaters (58%) said they had retested to improve their verbal scores, and a nearly

equal percentage (57%) had repeated the test to improve their quantitative scores. Fewer than a third of the test takers (32%) had retaken the test to increase their analytical scores. Undoubtedly, this lower percentage resulted from both the experimental nature of the analytical measure and the fact that some examinees had taken the GRE before 1977 when the analytical measure was introduced.

Nearly half (49%) of the responding test repeaters said they had repeated the test at the request of a graduate school. The current GRE program policy entails sending a cautionary note with any report that contains test scores that are more than five years old (Educational Testing Service, 1983a). We note that the percentage of candidates who retook the test at the request of a graduate school corresponds quite closely with the proportion of candidates whose most recent previous scores were at least five years old. Thus, it is not surprising to find that most persons who repeated the test within a few months or years reported that they did so to improve their scores, while most who repeated after several years reported that they needed more recent scores.

A sample of candidates' responses regarding why they retook the test gives some clues to candidates' motivations for retesting. By far the most frequent reasons mentioned were that previous scores had been obtained too long ago, were perceived as too low, or in some cases were both too old and too low. Often examinees retook the test because minimally acceptable score levels had been established by graduate departments, and the available scores did not meet these standards. Test takers often mentioned specific test score levels, usually a composite score on the verbal and quantitative sections. Examinees also frequently expressed the need to

update their test scores, either because graduate departments had explicitly requested they do so, or because they assumed that their scores were too old to be considered in admissions decisions. This assumption usually seemed to be based on general impressions, which were sometimes at least partially based on such information sources as graduate school catalogs.

Besides wishing to enhance their overall chances for admission to graduate school a significant number of candidates mentioned the role of GRE scores in determinations of fellowships, assistantships, and other forms of financial aid. Higher scores were often sought to increase the prospects of obtaining such awards.

A number of examinees, usually older or returning students, apparently retook the exams at least partially for self-evaluation purposes--out of "curiosity" to assess the effects of thought-to-be relevant intervening experiences (e.g., "to see if more education an experience improves scores"), and maturation (e.g., to "prove to myself that being older (57) may be slower, but not dumber)." Others stated that, although they remembered having taken the test before, either they had no recollection of their scores or else the earlier scores could not be found.

Finally, a few others wanted to update their credentials by taking the new analytical part of the test. Some seemed merely curious about how they would do on this new measure, while others thought the new measure might improve their chances for admission. A small number also admitted that their initial test was for practice only--merely a "trial run."

5. <u>What general activities do examinees engage in between test occasions? Are these activities related to test score improvements</u>?

To answer this question, examinees were asked to describe their between-test involvement in the following activities:

(a) further undergraduate study

(b) graduate study in each of several fields

(c) independent study, adult education, or group discussions

(d) employment experiences (and whether they involved extensive reading/writing, working with numbers, or problem solving), and

(e) other activities that may have affected their most recent test performance.

About 41% of examinees reportedly engaged in additional undergraduate study between test occasions; a plurality of these were in the social sciences. A majority (about 60%) also reported at least some involvement in employment that entailed extensive reading or writing. An approximately equal percentage said they had been engaged in employment that required the use of analytical or problem-solving skills. Somewhat fewer (49%) indicated employment experiences that involved numerical skills. For each of these employment categories, however, relatively few examinees reported "great involvement"--26%, 12%, and 19%, respectively.

Besides these activities, examinees also volunteered that they had been engaged in a wide variety of pursuits that they believed may have affected GRE test performance. These included such diverse activities as extensive travel, teaching/counseling, and community involvement. By far the most frequently mentioned activity was reading, usually as either a leisure time activity or a planned activity to increase particular skills (i.e., speed reading courses or vocabulary building exercises). None of the activities, however, or any indices based on composites of them, correlated significantly with score changes for any section of the test.

6. To what extent do test repeaters perceive various factors as affecting their test performances on initial testing and subsequent retesting? How do perceived effects relate to test score changes?

Examinees were asked about various factors that may have affected their test performances on each of their two most recent testings. Table 6 compares examinees' responses with respect to each test occasion.

Insert Table 6 about here

Table 7 summarizes these data, providing (a) examinees' average ratings for each factor and (b) the percentages of examinees who said they were affected more on one test occasion than another.

Insert Table 7 about here

From these two presentations, it appears that:

1. Lack of familiarity with specific types of question formats, lack of adequate review of subject matter, and slowness were perceived by examinees as being the most influential factors on both initial testing and retesting. There was a slight reordering of these "top three" on retesting, with "slowness" replacing "lack of familiarity" as the single most important factor.
2. No factor was seen as becoming more bothersome on retesting than on initial testing. The three factors whose perceived influence decreased most were all related to test

preparation--"lack of familiarity with general test-taking procedures," "lack of familiarity with specific types of question formats," and "lack of adequate review of test procedures."

3. Some factors were not perceived as very important on either the initial or subsequent test occasion. These were "sickness," "understanding general test directions," "personal problems," and "poor testing conditions." Although severe illness would probably adversely affect most test takers' performances, it appears that many students decide not to take the exam if they are sick. When asked about their reasons for not taking the test, even though they had registered, one of the most frequently given answers was "illness."

Examinees apparently have little trouble understanding general test directions and that they are bothered very little by what they consider to be poor testing conditions. In addition, with respect to being unlucky, examinees in our sample were slightly more likely to attribute any misfortune to the strategies they used to render guesses than to the luck of the draw in getting an unusually difficult test form, suggesting that alternate test forms are not perceived as differentially difficult.

Besides the factors listed on our questionnaire, examinees were encouraged to suggest other factors that may have affected their test performances on one or more occasions. Although most of the additional factors could be classified under those we had listed, a significant number of examinees also mentioned the time lapse between tests as a factor in score differences. Generally, these responses suggested that verbal

ability may have increased, while quantitative ability probably declined, due to either activity or inactivity in these domains. For instance, several test takers mentioned the general lack of a stimulating intellectual environment since college as a factor. Typical comments regarding expectations for higher verbal but lower quantitative scores were:

> "I haven't done any math for the 10 years between tests and really found it too boring to review much. I expect my verbal score to be higher now simply because of 10 more years on earth," and
>
> "[with regard to verbal activities], I've been doing crossword puzzles, reading magazines, . . . but math-wise I've done nothing beyond balancing my checkbook. . . ."

Correlations among examinees' assessments of the effects of various factors both across occasions and within occasions are shown in Table 8. (Correlations are among examinees' responses on a three-point scale.) Some factors (e.g., "sickness") appear to be essentially random, i.e., they were not perceived as being systematic influences across test occasions. For certain other factors, however, such as "being unlucky"--either in making guesses or in getting a difficult test form--examinees were quite consistent in their ratings for previous and recent tests. Ratings for these two "luck" factors correlate .71 and .70 across test occasions. Although "luck" was not rated overall as an especially important factor in test performances, examinees who felt that it was important apparently saw it as a consistent influence in testing, while those who perceived it as unimportant saw it as consistently so. Other personal traits like slowness, nervousness, and carelessness were also perceived as relatively consistent influences, with correlations of .62, .54, and .53 across occasions, respectively.

Insert Table 8 about here

The relatively high correlations among Factors A, B, C, M, and N suggest the presence of a "test preparation" factor, which was stronger on the previous test occasion than on the more recent testing. The median correlation among ratings of these factors was higher (.43) for the previous testing than for the more recent testing (.35), and each of the corresponding correlations was lower for the retest than for the previous test. We note that repeaters in this study were generally more actively engaged in test preparation than were GRE test takers in general. When compared with examinees from the same test administration who had not repeated the test (Powers & Swinton, 1982, p. 8), test repeaters spent about 20% more time preparing for each section of the test.

When asked to suggest for each test section the single most important explanation of any score differences between previous and most recent testings, examinees tended to mention one of the four factors listed in Table 9. For both the verbal and quantitative test sections, the lack of adequate subject matter preparation was most often mentioned as the most important factor. Nearly a third of the sample mentioned this for the quantitative section. For the analytical section, however, the lack of familiarity with question formats was by far the most often mentioned reason for test score changes. This factor was rated as the second most important for each of the other two test sections. Again it should be noted here that in 1980, when these data were collected, the GRE analytical ability section contained four item types, two of which have since been removed because of their susceptibility to coaching and practice. This susceptibility was thought to stem in large part from the complexity of the item types.

Insert Table 9 about here

"Slowness" was seen as the third most important factor for each section of the test, being perceived as somewhat more important for the verbal section than for either the quantitative or the analytical sections. In this regard, we noted that a significant number of examinees mentioned their inability to read rapidly as a contributor to low verbal scores.

Finally, tiredness was mentioned by about one in every 10 test takers as the most important explanation of test score differences for each section of the test. Examinees sometimes mentioned such factors as (a) work commitments or personal problems that prevented a good night's sleep before the test, (b) the early hour of the test administration (which was viewed as penalizing those who were not "morning people"), (c) the physical and mental fatigue resulting from a heavy academic load, and (d) the fatigue caused by the test itself ("You get tired and lose concentration after a period of time" and "Three hours is a long time to sit in a room and make 'letter dots'!")

To evaluate the degree to which changes over test occasions in each of these factors were related to test score changes, a difference index was computed for each examinee on each factor. This index was correlated with test score changes for each section of the test. By and large, the relationships were small and nonsignificant. The largest correlations were between a composite variable reflecting a reduction in the lack of preparedness upon retesting and test score changes on the quantitative (.17) and the analytical (.19) sections of the test. This composite

reflected the total differences in test preparation factors A, B, C, M, and N. The only other significant correlations were between changes in quantitative scores and composite variables reflecting reductions in the perceived effects of (a) personal factors (sickness, carelessness, tiredness, personal problems, nervousness, and slowness) ($r = .14$, $p < .01$) and (b) external factors (being unlucky and poor testing conditions) ($r = .14$, $p < .01$). For personal factors, tiredness was the factor that contributed most to the correlation with the composite index.

7. What other comments do examinees have about possible reasons for test score changes?

When asked to volunteer any other information they thought relevant to test score changes, examinees most often mentioned differences in being prepared to take the tests. By "being prepared" examinees frequently meant familiarity with the test ("I improved simply because I was more familiar with the test format.)" At least equally often, however, it also referred to a more general lack of preparedness:

> "The first time, I was nervous, tired, and totally unprepared. This time I was better prepared both mentally and emotionally."
>
> "School work and tests are simply [a matter of developing] a mind set, a pattern of thinking. A return to school helped in this regard," and
>
> "I was much more prepared, both physically and psychologically for the recent exam."

The relationship between test familiarity and the more general state of preparedness was expressed by one test taker:

> "I had a better mental disposition this time because of increased familiarity with what to expect on the GRE."

Although we did not attempt a systematic assessment, there appeared to be a difference in test readiness between older, returning test takers and those who had completed their undergraduate degrees more recently. The relatively recent GRE program emphasis in providing more extensive test familiarization materials may have helped some older students with respect to recent testing. As one examinee lamented,

> "I didn't have any idea of what I was taking in 1970. I don't even remember seeing an information bulletin."

But the time lapse between academic activities and admissions testing may have hurt some examinees:

> From one, "I've been out of school five years and test-taking skills have declined,"

and from another

> "It was easier to take the test the first time at the end of college when test-taking was routine."

Returning test takers, particularly those for whom the time between tests was significant, also frequently mentioned their years of non-use of mathematics as a reason for declines in quantitative test scores. Most often test takers merely cited their rusty quantitative skills, but a few felt that because they placed more emphasis on the "new math" and the metric system, recent test versions clearly favored younger examinees. Typical of the comments were:

> "My mid-level management position does not lend itself to many quantitative experiences." and
>
> "It's been 25 years since I had any math, and as an RN, I do not use algebra or geometry. My math scores will be considerably lower. . ."

Verbal experiences were much more likely to be seen as a positive influence on verbal scores:

> "Several years of experience as an editor and journalist should make a substantial difference in my verbal scores."

Returning test takers frequently mentioned such personal qualities as increased maturity and emotional stability as factors in improved test performance:

> "If my scores are higher, it's due to [additional academic work] and 12 years of experience and maturation."
>
> "I'm more settled down now that I'm out of college" and
>
> "I'm 8 years older . . . , more mature, and a lot calmer."

For many examinees, this maturity seemed to be accompanied by a greater sense of purpose and perspective. Increased motivation--either self-induced or the result of specific graduate admissions requirements--was a frequently mentioned factor in improved test performance. Some typical comments were:

> "The first time I took the test I was not applying to graduate school and had no interest in graduate education. Now, I want to get into a specific school, so I am more motivated to do well,"

and

> "The first time [I took the test] I had just completed an undergraduate degree. I was tired, burned out, and didn't care what my score was. I was fed up with school and with tests. Not so now!"

Some examinees apparently were able to take the exams more seriously upon retesting:

> "I think the realization of the seriousness and importance of making a higher score was a factor [in improved performance]"

Others were better able to put the test in proper perspective:

> "As an undergraduate these tests had greater significance relative to my future. Their significance and the pressure around them was reduced greatly from the perspective of my age."

Throughout test takers' comments, one of the most frequently mentioned factors was time since either the receipt of an undergraduate degree or previous testing, or simply age. Perhaps one reason for the difficulty in being able to specify any of the factors as actual causes of test score

gains or losses is that different examinees sometimes gave the same explanation for gains as others gave for decreases. For example, the better perspective and increased motivation that came with age, was also seen as being accompanied by a decay in general test taking skills or in specific quantitative skills. For some examinees, the time between tests apparently enabled significant reading and personal development that was viewed as a positive influence on test scores; other test takers described the same period of time as involving so little intellectual stimulation that no improvement in abilities or test scores could be reasonably expected.

Discussion

We began by asking what additional advice might be offered to (a) test takers contemplating whether or not to repeat a test like the GRE General Test and (b) test users confronted with how to treat multiple test scores in the admissions process. The advice currently offered in test program publications to prospective GRE General test takers is that

> "Unless your scores seem unusually low compared with other indicators of your ability, taking the GRE again is not likely to result in a substantial score increase." (Educational Testing Service, 1983a, p. 54)

This study provides no startling new information that would indicate a need to modify this advice. The study does, however, provide additional support for other advice conveyed to test takers--namely, the desirability of preparing for the test. Currently, GRE test takers are advised that, by taking the sample practice test that is included in the Bulletin, they may achieve the benefit of a test practice effect without having to repeat the actual test. Our results suggest indirectly that examinees who are best

able to judge, i.e., test repeaters, would probably endorse this advice. Perhaps if included in test information bulletins, such testimony from ~~fellow test takers would provide~~ further incentive for examinees to prepare for tests like the GRE General Test. For the GRE, examinees could now be told that

> "One reason that GRE test takers often give for test score changes from one occasion to another is the lack of appropriate preparation for the initial test. This suggests the desirability of gaining some pre-examination familiarity with the test, for example, by using this information bulletin or other appropriate test preparation resources."

The results of this study also have implications for test users. Those who use GRE test scores are currently told of the "retest effect"--that, on average, test repeaters show a score gain of 25-30 points on the GRE General Test and that repeaters seem to be a self-selected group of test takers, who have lower-than-average initial test scores. Several possible ways of evaluating multiple test scores are mentioned, including the use of (a) the highest score, (b) the most recent one, or (c) the average of all test scores. It is suggested that the average of several scores, if earned in a short period of time, may be the best technique, but that whatever approach is adopted, it should be used consistently with all applicants (Educational Testing Service, 1983b).

Our results do not suggest a need for any radical alteration of this advice, which seems sound. Similar advice has been offered for the Law School Admission Test (LSAT) after extensive analyses of repeater test scores by Linn (1977) who found that no single method of treating multiple LSAT scores had proven clearly superior (although using only the initial score had proven clearly inferior). Linn (1977) therefore recommended using the average of scores earned on all occasions, except when there is a

good reason to discount one of the scores. Our study suggests that the most plausible reason for discounting the validity of initial test scores is the lack of preparation for the test, since, in the eyes of examinees, this was the factor that most often adversely affected initial test performance. Our results suggest that the lack of preparedness is something that examinees apparently feel they can overcome upon retesting. Being better prepared may reduce the extraneous variation in test scores due to facility with test-taking procedures; if so retest scores might reasonably be expected to yield more valid predictions of future academic performance than would initial scores. The same line of reasoning would seem to apply to such factors as slowness and fatigue.

GRE test users are also warned to exercise caution in interpreting score gains as an indication of academic development (Educational Testing Service, 1983b). Our results reinforce this admonition: although examinees often felt that certain between-test experiences contributed to either higher or lower subsequent test scores, we were unable to document any consistent relationships between these activities and test score changes. This finding suggests that if such relationships do exist, then better methods are needed to document relevant intervening activities, and to index changes in developed abilities.

Finally, one general finding was that examinees have definite opinions (which they are willing to volunteer) about discrepancies between multiple sets of test scores, and these opinions are systematically though rather weakly related to test performance. Our advice to test takers would be to make these opinions known to admissions committees. To test users we suggest that, when confronted with discrepant multiple test scores, they

consider seeking test takers' opinions about reasons for score changes. This exchange might have potential not only for explaining test score changes, but also for providing other interesting information about applicants that might not otherwise be gleaned from personal statements or application forms.

Since this study was initiated, the GRE General Test has undergone two significant changes that are quite consistent with the results of this study. First, the analytical portion of the test has been revised so that it no longer contains two item types that were found to be very susceptible to improvement through practice and test preparation. If we were to repeat this study, we would expect to find somewhat less concern about test preparation as a factor in test score changes. Second, the number of questions in the verbal portion of the test has been reduced and the time allotted for this section has been increased. We would also expect, therefore, that examinees' would perceive the influence of test speededness (and their slowness) as being less influential for the revised verbal measure. It seems likely that all of the other study conclusions would pertain to the revised GRE General Test and, although we cannot be absolutely certain, it is also probable that many of the conclusions would also apply to examinees who retake other major admissions tests.

References

Alderman, D. L. (1981a). Student self-selection and test repetition. Educational and Psychological Measurement, 41, 1073-1081.

Alderman, D. L. (1981b). Measurement error and SAT score change (College Board Report No 81-9; also RR No. 81-39). Princeton, N.J.: Educational Testing Service.

Boldt, R. F. (1977). Evaluation of three methods for treating repeaters' scores. Report #LSAC-77-7. In Law School Admission Council, Reports of LSAC sponsored research (Vol. 3, 1975-77). Princeton, N.J.: Law School Admission Council.

Campbell, J. T., Hilton, T. L., & Pitcher, B. (1967). Effects of repeating on test scores of the Graduate Record Examinations (GRE Special Report No. 67-1). Princeton, N.J.: Educational Testing Service.

Clark, M. J. (1984). Older and younger graduate students: A comparison of goals, grades, and GRE scores (GRE Board Research Report No. 81-17R). Princeton, N.J.: Educational Testing Service.

Educational Testing Service. (1983a). Graduate Record Examinations 1983-84 information bulletin: Registration preparing for the test publications and registration for minority graduate student locatei service. Princeton, N.J.: Educational Testing Service.

Educational Testing Service. (1983b). GRE guide to the Graduate Record Examinations Program 1983-84. Princeton, N.J.: Educational Testing Service.

Goodison, M. B. (1982). A summary of data collected from Graduate Record Examinations test-takers during 1980-81 (Data Summary Report #6). Princeton, N.J.: Educational Testing Service.

Goodison, M. B. (1983). A summary of data collected from Graduate Record Examinations test-takers during 1981-82 (Data Summary Report #7). Princeton, N.J.: Educational Testing Service.

Hartle, T., Baratz, J., & Clark, M. J. (1983). Older students and the GRE Aptitude Test (GRE Board Research Report No. 76-13R). Princeton, N.J.: Educational Testing Service, 1983.

Jacobs, P. I. (1966). Large score changes on the Scholastic Aptitude Test. Personnel and Guidance Journal, 150-156.

Kingston, N., & Turner, N. (1984). Analysis of score change patterns of examinees repeating the Graduate Record Examinations General Test (Draft final report GRE No. 83-5 to the GRE Research Committee). Princeton, N.J.: Educational Testing Service.

Leary, L. F., & Wightman, L. E. (1983). Estimating the relationship between use of test-preparation methods and scores on the Graduate Management Admission Test (GMAC Research Report 83-1; also ETS RR-83-22). Princeton, N.J.: Educational Testing Service.

Levine, R. S., & Angoff, W. H. (1958). The effects of practice and growth on scores on the Scholastic Aptitude Test. Statistical Report 58-6. Princeton, N.J.: Educational Testing Service.

Linn, R. L. (1977). On the treatment of multiple scores for Law School Admission Test repeaters. Report #LSAC-77-7. In Law School Admission Council, Reports of LSAC sponsored research (Vol. 3, 1975-77). Princeton, N.J.: Law School Admission Council.

Olsen, M., & Schrader, W. B. (1959). The use of preliminary and final Scholastic Aptitude Test scores in predicting college grades (College Entrance Examination Board Research and Development Reports; also Statistical Report 59-19). Princeton, N.J.: Educational Testing Service.

Pitcher, B. (1966). Test-taking patterns of College Board candidates tested in 1962-1964 based on a five percent sample (Statistical Report 66-72). Princeton, N.J.: Educational Testing Service.

Powers, D. E., & Swinton, S. S. (1982). Effects of self-study of test familiarization materials for the analytical section of the GRE Aptitude Test (GRE Board Research Report No. 79-9R). Princeton, N.J.: Educational Testing Service.

Rock, D., & Werts, C. (1980). An analysis of time-related score increments and/or decrements for GRE repeaters across ability and sex groups (GRE Board Research Report No. 77-9R). Princeton, N.J.: Educational Testing Service.

Stricker, L. J. (1982). Test disclosure and retest performance on the Scholastic Aptitude Test (College Board Report No. 82-7; also RR-82-48). Princeton, N.J.: Educational Testing Service.

Watkins, R. W., & Schrader, W. B. (1963). The relation of differences between junior and senior Scholastic Aptitude Test scores and college grades (College Entrance Examinations Board Research and Development Reports, and Statistical Report 63-29). Princeton, N.J.: Educational Testing Service.

Wild, C. L. (1981). A summary of data collected from Graduate Record Examinations test-takers during 1979-80 (Data Summary Report #5). Princeton, N.J.: Educational Testing Service.

Table 1

Number of Cases Available for Analysis

	Questionnaire		
Two Test Scores	Yes	No	Total
Yes	433	147	580
No	195	89	284
Total	628	236	864

Table 2

Comparison of Study Sample with All GRE Test Takers

Variable		All 1979-80 GRE Test Takers[a] (N ≈ 210,000)	Primary Study Sample (N = 433)
Undergraduate Field (%)			
Humanities		15.4	14.0
Social Sciences		42.9	48.8
Biological Sciences		31.8	27.0
Physical Sciences		17.1	10.2
Sex (% female)		53.6	51.0
Racial Identification			
Black		6.7	21.1
White		86.3	79.0
Other		7.0	0.0
Degree Objective			
Nondegree		0.9	0.2
Master's or intermediate		62.7	46.6
Doctorate or postdoctoral study		36.3	53.1
GRE Scores			Previous / Recent
Verbal	M	487	453 / 484
	SD	123	118 / 127
Quantitative	M	516	460 / 478
	SD	131	131 / 131
Analytical[b]	M	508	438 / 500
	SD	127	124 / 130
English Best Language (%)		92.4	95.4
Year of Receipt of Bachelor's Degree			
1969 or earlier		7.8	14.7
1970 - 1974		11.6	26.4
1975 - 1979		36.0	45.7
1980 or later		42.7	13.3

[a]Source: Wild, C. L. A summary of data collected from Graduate Record Examinations test-takers during 1979-80. Data Summary Report #3. Princeton, N.J.: Educational Testing Service, 1981.

[b]Based on analytical scores earned before the analytical measure was revised in October 1981.

Table 3

Time Between June Testing and Most Recent Previous Testing

	Percent of Examinees	
Time Interval	Reported (N = 628)	Test Files (N = 433)
Less than 6 months	11.3	12.0
6 months to 2 years	32.7	36.0
3 to 8 years	33.5	38.4
9 years or more	22.5	13.6

Table 4

Patterns of GRE Score Differences
(Second minus First Score)

	Section		
Statistic	Verbal (N = 433)	Quantitative (N = 433)	Analytical (N = 229)
Median	23.3	8.6	47.1
Mean	31.1	18.5	56.7
SD	57.7	70.6	67.3
% Decreases:			
Total	26.1	35.6	17.5
50 points or more	8.5	16.9	5.2
% Increases:			
Total	67.9	56.4	77.3
100 points or more	13.9	13.2	29.3
% No Change	6.0	8.1	5.2

Note. Fewer test takers repeated the analytical section of the test because it was first offered in October 1977.

Table 5

Patterns of Average GRE Scores by Length of Time Between Tests

Test Score		Time Between Tests: Less than 6 months (N=57)	6 months–2 years (N=168)	3 – 8 years (N=155)	9 years or more (N =53)
Verbal	June 1980	431	450	521	536
	Previous	413	422	488	491
	Difference	18	28	33	45
Quantitative	June 1980	451	475	476	492
	Previous	437	444	470	482
	Difference	14	31	06	10
Analytical	June 1980	480	500		
	Previous	435	439		
	Difference	45	61		

Note. The analytical section was first given in 1977; therefore, very few members of the sample were repeating the test after three years or more.

Table 6

Examinee Ratings of Test Performance Factors on Two Test Occasions

Factor	Test Occasion	Effect (Percentages of Examinees) Little or No Effect	Some Effect	Great Effect
Lack of familiarity with general test-taking procedures	Previous	48.1	32.5	19.4
	Recent	77.7	18.5	3.8
Lack of familiarity with specific types of question formats	Previous	29.5	42.9	27.6
	Recent	56.4	32.9	10.7
Not understanding test directions	Previous	76.8	16.8	6.4
	Recent	87.8	10.7	1.5
Being unlucky in making guesses	Previous	55.1	31.1	13.9
	Recent	66.1	28.2	5.7
Being unlucky in getting a form of the test with unexpected questions	Previous	70.2	18.8	11.0
	Recent	72.2	21.1	6.7
Sickness	Previous	87.9	7.5	4.5
	Recent	92.5	5.5	2.0
Carelessness	Previous	54.6	36.8	8.6
	Recent	68.2	30.0	1.8
Tiredness	Previous	57.3	26.9	15.8
	Recent	65.0	26.8	8.2
Personal problems	Previous	73.1	18.0	8.9
	Recent	80.3	16.1	3.6
Nervousness	Previous	46.6	37.3	16.1
	Recent	57.6	34.3	8.1
Slowness	Previous	39.0	36.6	24.4
	Recent	40.7	42.5	16.8
Poor testing conditions	Previous	80.7	14.5	4.8
	Recent	79.2	15.5	5.3
Lack of adequate review of test procedures	Previous	49.3	30.6	20.1
	Recent	74.6	20.2	5.2
Lack of adequate review of subject matter	Previous	33.2	36.4	30.4
	Recent	45.8	38.6	15.6
Other	Previous	91.2	2.0	6.8
	Recent	90.1	3.0	6.6

Note. All percentages are based on between 603 and 613 respondents.

Table 7
Examinee Perceptions of Effects of Various Factors on Test Performance

Factor	Mean (SD)[a] Previous Test	Recent Test	Effect Size[b]	Percentages of Examiness who Reported: More Effect on Recent Test	Some Effect on Both Tests	More Effect on Previous Test
Lack of familiarity with general test-taking procedures	1.70 (.77)	1.26 (.52)	.57	4.6	53.7	41.7
Lack of familiarity with specific types of question formats	1.98 (.76)	1.55 (.68)	.57	12.0	40.8	47.2
Not understanding test directions	1.29 (.58)	1.14 (.38)	.26	4.3	79.1	16.6
Being unlucky in making guesses	1.58 (.72)	1.40 (.60)	.25	2.3	78.5	19.2
Being unlucky in getting a form of the test with unexpected questions	1.40 (.68)	1.35 (.60)	.07	5.8	83.1	11.2
Sickness	1.17 (.48)	1.09 (.35)	.25	5.8	82.9	11.3
Carelessness	1.54 (.65)	1.33 (.51)	.32	4.3	73.5	22.2
Tiredness	1.59 (.75)	1.43 (.64)	.21	14.6	61.1	24.3
Personal problems	1.37 (.65)	1.23 (.50)	.22	9.0	73.7	17.2
Nervousness	1.69 (.73)	1.50 (.64)	.26	8.7	65.5	25.8
Slowness	1.85 (.78)	1.76 (.72)	.12	12.4	67.7	19.9
Poor testing conditions	1.24 (.53)	1.26 (.55)	-.04	12.5	76.9	10.5
Lack of adequate review of test procedures	1.71 (.78)	1.31 (.56)	.51	5.9	58.4	35.7
Lack of adequate review of subject matter	1.97 (.80)	1.70 (.72)	.34	12.3	54.5	33.2
Other	1.16 (.53)	1.17 (.53)	-.02	3.6	93.2	3.1

[a] Responses were on a three point scale with 1 = little or no effect on test performance, 2 = some effect, and 3 = great effect.

[b] Effect size is the difference between previous and recent tests divided by the SD of the ratings for the previous test.

Note. With the exception of the "other" factor, all numbers are based on the responses of between 603 and 613 examinees.

Table 8

Correlations among Factors Affecting Test Performance

Previous test	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
A. Lack of familiarity with general test-taking procedures		56	37	31	26	11	20	13	22	30	35	11	46	38	-01	39	05	08	21	17	03	10	01	06	15	16	03	15	09	-08
B. Lack of familiarity with specific types of question formats			39	29	25	07	22	20	22	26	28	13	44	41	07	20	23	07	23	13	01	12	05	06	14	11	04	13	11	-02
C. Not understanding test directions				26	20	06	25	14	20	24	26	09	39	26	05	14	03	39	17	15	00	09	01	05	08	10	02	13	06	-02
D. Being unlucky in making guesses					48	09	34	20	17	27	27	17	29	31	03	17	09	14	71	33	09	21	15	12	20	14	12	16	17	-00
E. Being unlucky in getting a form of the test with unexpected questions						10	25	19	24	28	28	23	27	26	10	16	13	08	37	70	12	16	08	14	17	16	11	13	15	05
F. Sickness							23	39	26	15	07	13	16	17	-04	03	-07	02	04	04	05	06	02	08	06	01	01	03	00	-04
G. Carelessness								32	30	34	32	19	32	32	06	06	-12	03	19	10	11	53	08	08	20	14	04	09	07	02
H. Tiredness									38	24	21	19	24	20	04	-01	-02	02	10	02	00	13	28	12	14	08	05	04	01	04
I. Personal problems										35	28	18	27	24	03	11	07	05	05	08	01	15	11	33	20	14	08	10	05	-03
J. Nervousness											43	25	27	28	07	09	03	07	15	18	04	16	12	12	54	22	07	12	14	-01
K. Slowness												19	26	33	02	14	01	11	19	22	06	16	08	17	23	62	09	11	16	-01
L. Poor testing conditions													14	15	15	09	06	08	14	15	05	17	10	09	15	10	27	12	16	10
M. Lack of adequate review of test procedures														51	05	12	-04	06	14	12	-03	15	-00	-05	11	10	01	36	12	00
N. Lack of adequate review of subject matter															09	08	-00	05	22	19	05	12	07	10	15	15	03	15	38	-05
O. Other																-01	-01	06	-04	09	-03	02	-00	-01	-03	-01	-02	-02	-03	68
Recent test																														
A. Lack of familiarity with general test-taking procedures																45	27	20	23	09	11	06	12	20	24	02	35	24	05
B. Lack of familiarity with specific types of question formats																	38	20	21	07	06	21	19	18	19	04	36	33	02
C. Not understanding test directions																		16	18	16	07	10	13	10	14	07	33	18	09
D. Being unlucky in making guesses																			41	16	23	23	16	24	20	18	18	27	-03
E. Being unlucky in getting a form of the test with unexpected questions																				14	13	10	15	18	23	11	19	22	05
F. Sickness																					14	27	16	14	10	14	06	10	-02
G. Carelessness																						23	14	21	22	08	17	20	06
H. Tiredness																							41	28	14	16	18	22	03
I. Personal problems																								34	23	15	14	23	05
J. Nervousness																									33	16	17	27	03
K. Slowness																										13	17	32	10
L. Poor testing conditions																											03	08	00
M. Lack of adequate review of test procedures																												48	03
N. Lack of adequate review of subject matter																													04
O. Other																													

Note. Correlations are based on 627 examinees who took the GRE Aptitude Test two times and who returned questionnaires. Decimals have been omitted.

Table 9

Examinees' Perceptions of Factors Most Related to Score Changes

	Test Section		
Factor	Verbal	Quantitative	Analytical
Lack of adequate review of subject matter	22.7%	33.1%	16.6%
Lack of familiarity with specific question formats	16.0	14.8	29.1
Slowness	15.5	11.6	11.3
Tiredness	10.4	10.1	10.5

Note. No other factor was nominated by more than 10% of examinees for any section of the test. Percentages are based on 595 respondents for the verbal and quantitative sections and 506 for the analytical section.