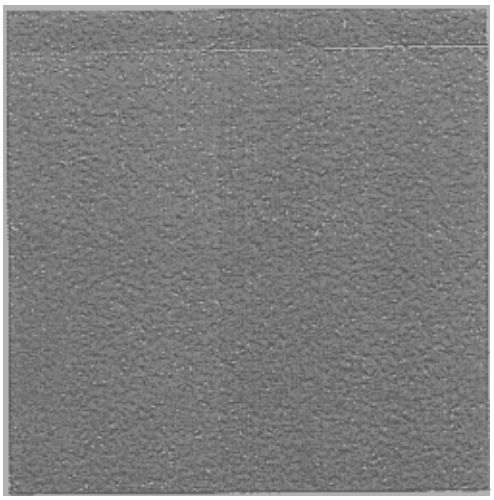

EFFECT OF INCREASED TEST-TAKING TIME ON TEST SCORES BY ETHNIC GROUP, AGE, AND SEX

Cheryl Wild and Robin Durso

GRE Board Research Report GREB No. 76-6R

June 1979

This report presents the findings of a research project funded by and carried out under the auspices of the Graduate Record Examinations Board.

EDUCATIONAL TESTING SERVICE, PRINCETON, NJ

GRE BOARD RESEARCH REPORTS
FOR GENERAL AUDIENCE

Altman, R. A. and Wallmark, M. M. A Summary of Data from the Graduate Programs and Admissions Manual. GRE Board Research Report GREB No. 74-1R, January 1975.

Baird, L. L. An Inventory of Documented Accomplishments. GRE Board Research Report GREB No. 77-3R, June 1979.

Baird, L. L. Cooperative Student Survey (The Graduates [$2.50 each], and Careers and Curricula). GRE Board Research Report GREB No. 70-4R, March 1973.

Burns, R. L. Graduate Admissions and Fellowship Selection Policies and Procedures (Part I and II). GRE Board Research Report GREB No. 69-5R, July 1970.

Centra, J. A. How Universities Evaluate Faculty Performance: A Survey of Department Heads. GRE Board Research Report GREB No. 75-5bR, July 1977. ($1.50 each)

Centra, J. A. Women, Men and the Doctorate. GRE Board Research Report GREB No. 71-10R, September, 1974. ($3.50 each)

Clark, M. J. The Assessment of Quality in Ph.D. Programs: A Preliminary Report on Judgments by Graduate Deans. GRE Board Research Report GREB No. 72-7aR, October 1974.

Clark, M. J. Conditions Influencing the Career Accomplishments of Recent Ph.D.'s. GRE Board Research Report GREB No. 76-2R, June 1979.

Clark, M. J. Program Review Practices of University Departments. GRE Board Research Report GREB No. 75-5aR, July 1977. ($1.00 each)

Donlon, T. F. Annotated Bibliography of Test Speededness. June 1979.

Flaugher, R. L. The New Definitions of Test Fairness In Selection: Developments and Implications. GRE Board Research Report GREB No. 72-4R, May 1974.

Fortna, R. O. Annotated Bibliography of the Graduate Record Examinations. July 1979.

Frederiksen, N. and Ward, W. C. Measures for the Study of Creativity in Scientific Problem-Solving. May 1978.

Hartnett, R. T. and Willingham, W. W. The Criterion Problem: What Measure of Success in Graduate Education? GRE Board Research Report GREB No. 77-4R, March 1979.

Knapp, J. and Hamilton, I. B. The Effect of Nonstandard Undergraduate Assessment and Reporting Practices on the Graduate School Admissions Process. GRE Board Research Report GREB No. 76-14R, July 1978.

Lannholm, G. V. and Parry, M. E. Programs for Disadvantaged Students in Graduate Schools. GRE Board Research Report GREB No. 69-1R, January 1970.

Reilly, R. R. Critical Incidents of Graduate Student Performance. GRE Board Research Report GREB No. 70-5R, June 1974.

Rock, D. A. The Prediction of Doctorate Attainment in Psychology, Mathematics and Chemistry. GRE Board Research Report GREB No. 69-6aR, June 1974.

Schrader, W. B. Admissions Test Scores as Predictors of Career Achievement in Psychology. GRE Board Research Report GREB No. 76-1R, September 1978.

Wild, C. L. Summary of Research on Restructuring the Graduate Record Examinations Aptitude Test. February 1979.

Wilson, K. M. The GRE Cooperative Validity Studies Project. GRE Board Research Report GREB No. 75-8R, June 1979.

Wiltsey, R. G. Doctoral Use of Foreign Languages: A Survey. GRE Board Research Report GREB No. 70-14R, 1972. (Highlights $1.00, Part I $2.00, Part II $1.50).

Witkin, H. A.; Moore, C. A.; Oltman, P. K.; Goodenough, D. R.; Friedman, F.; and Owen, D. R. A Longitudinal Study of the Role of Cognitive Styles in Academic Evolution During the College Years. GRE Board Research Report GREB No. 76-10R, February 1977 ($5.00 each).

Effect of Increased Test-Taking Time on Test Scores by Ethnic Group, Age, and Sex

Cheryl Wild and Robin Durso

GRE Board Research Report GREB No. 76-6R

June 1979

Abstract

This study investigates the effects of increasing the test time to reduce the speededness of the verbal and quantitative experimental sections of the Graduate Record Examinations (GRE) Aptitude Test. In December 1976, at approximately 550 domestic test centers, 20- and 30-minute versions of a verbal experimental test and of a quantitative experimental test were administered with the operational test. The effects of testing time on scores for minority-group versus majority-group examinees, on male versus female examinees, and on examinees by years elapsed since they received the baccalaureate degree are investigated. Although the 20-minute experimental tests are generally more speeded than the 30-minute tests, the ten additional minutes resulted in a small score gain for all groups, and differential score gains were not found between the subgroups. On the basis of the results of this study, the extension of testing time for the GRE Aptitude Test for the purpose of reducing intergroup differences is not indicated.

Effect of Increased Test-taking Time on Test Scores by Ethnic Group, Age, and Sex

A great deal of concern has been expressed in recent years about the fairness of standardized aptitude tests. Unfairness can be introduced by test content, test administration procedures, and the relationship of the test to success. The possible "biasing agent" investigated in this study was test-taking time. The GRE Aptitude Test is designed to be a measure of "power" rather than "speed"; that is, it is expected that most candidates will have time to complete a majority of the questions within the time allotted. If certain subgroups of examinees are systematically given too little time to complete the test sections, it may be that their scores are depressed because of timing limitations and not because of less ability.

Although several studies have investigated the effect of test-taking time on the scores of minority groups, little evidence showing subgroup bias has been found. In two studies of reading comprehension questions on the Law School Admissions Test, Evans and Reilly (1972a, 1972b) found that increasing the amount of time per question did not differentially increase scores; i.e., although scores might improve with increasing time, the increase did not vary with race (1972a, 1972b) or sex (1972a). A similar study utilizing a special quantitative section of the Admission Test for Graduate Study in Business again reported no significant interaction between ethnic groups and time factors (Evans & Reilly, 1973). Although the test appeared to be more speeded for the members of the Black group, the extra time was not differentially beneficial to them.

Although no "biasing" effects resulting from timing were found in the previously cited studies, the results may not be directly comparable to those of the GRE Aptitude Test. The GRE test-taking population is different from those of the experimental studies by Evans and Reilly. In addition, the effects of timing may depend on the specific timing and the test content, which vary by testing program. Only one of the studies considered sex as a possible factor, and none considered years since the completion of the undergraduate degree.

The purpose of this study was to investigate the effect of timing for various subgroups of examinees on GRE Aptitude verbal and quantitative tests. For subgroups defined by sex, race, and years since receipt of the baccalaureate degree the effects on verbal and quantitative aptitude measures were investigated. Since timing might affect scores, the timing of experimental verbal and quantitative sections was manipulated rather than the timing of the

---

The authors would like to thank Donald Rubin for consultation on data analysis procedures, Richard Harrison for programming the special analyses, and Ruth Mroczka for running the item analyses.

operational sections, the scores of which are used in reporting. The following research question was investigated separately for verbal and quantitative experimental sections of the GRE Aptitude Test: Does increasing the amount of time per question have a differential effect on the scores (after controlling for initial ability) of examinee subgroups as defined by sex, by race (Black and White), and by the number of years that have elapsed since the baccalaureate degree was obtained?

Procedures

In the following sections, procedures used in conducting the study are described. In the first section a description of the experimental and operational GRE tests is presented. Next, the subject groups are described and are compared with a national GRE test-taking group. Background data used in the analyses are described. Finally, the two methods of analysis, item analysis and regression analysis, are explained.

Test Content

GRE Aptitude Tests prior to October 1977 contained 95 verbal questions and 55 quantitative questions. The verbal score was based on two separately timed sections, a 25-minute section of 55 discrete questions (27 seconds per question) and a 50-minute section of 40 reading comprehension questions (75 seconds per question). The operational quantitative section contained 55 questions presented in 75 minutes (about 82 seconds per question). Along with the three sections that contribute to an examinee's scores, a fourth "experimental" section was also administered. The scores on the experimental sections did not contribute to the scores received on the operational sections of the GRE Aptitude Test.

Two tests, one containing verbal questions and the other quantitative questions, were developed to be administered in the experimental section of the GRE Aptitude Test and to parallel the operational verbal and quantitative sections of the test. The experimental verbal test consisted of 26 questions: 11 reading comprehension questions (one short passage, one long) and 15 discrete questions. The experimental quantitative test consisted of 14 questions: 10 discrete mathematics questions and 4 data interpretation questions. Each experimental test was administered with a 20-minute and a 30-minute timing. The 20-minute verbal test allowed approximately 46 seconds per question (comparable to the operational verbal test in terms of time per question), and the 30-minute verbal test allowed approximately 69 seconds per question. The two versions of this test were identical; only the timing differed. The 20-minute quantitative test allowed approximately 86 seconds per question (comparable to the operational quantitative test in terms of time per question). The 30-minute quantitative test allowed approximately 129 seconds per question.

Each GRE Aptitude Test book contained the three operational sections and one section having a version (20 or 30 minutes/verbal or quantitative) of the experimental test. All sections were formula scored (rights minus one-fourth of the wrongs). Only one timing (20-minutes or 30-minutes) could be administered at a single test center. The timing (20 or 30 minute) was assigned randomly to test centers, and test center supervisors were sent specific instructions concerning the appropriate timing. Within test centers, verbal and quantitative experimental sections were assigned alternately to examinees. The dependent variable was the formula score on the experimental test. The covariate was the formula score on the corresponding operational sections (i.e., the two verbal sections or the one quantitative section).

Subjects

The experimental tests were administered at the December 1976 GRE national administration, which included candidates from approximately 550 domestic testing centers. Average scores of examinees who take the GRE in the December administration tend to be slightly higher than the average scores of examinees at all administrations during the year. For all examinees during the 1976-77 testing year, the mean verbal and quantitative scores were 490 and 514, respectively. The mean verbal and quantitative scores for all December 1976 examinees taking an experimental test were 513 and 523, respectively. Thus, tests may appear to be slightly less speeded for the December group than for the total test-taking population.

Background Information

The following background information was collected by means of optional questions on the GRE registration form: (1) the number of years since the candidate obtained the baccalaureate degree; (2) the sex of the candidate; and (3) the ethnic group of the candidate. A statement in the Information Bulletin assures examinees that their responses to these questions will be kept confidential and used for research purposes only. Examinees who did not complete one or more of these questions were not included in the analysis. The number of years since receipt of the baccalaureate degree was divided into five categories: 0-4 years, 5-9 years, 10-14 years, 15-19 years, and 20-24 years. Black and White ethnic groups were studied.

Data Analysis

Two kinds of procedure were used to analyze the data, regression analysis and item analysis. Regression analysis is used in this study to determine whether increasing the amount of time per question has a differential effect on the scores of specific examinee groups

other than the differential effect expected by initial score differences. The item analysis was done for descriptive information only, and no control for initial score difference is done with this technique. Since item analysis techniques are traditionally used at ETS to examine test speededness, it was decided that a comparison of item analysis and regression analysis results would be useful both in terms of explicating regression analysis results and in interpreting future item analysis results.

Regression analysis for estimating timing effects. Regression analysis was used to estimate the effects of timing after controlling for initial score differences. The regression analyses are essentially analyses of covariance in randomized experiments. However, since the assignment of experimental test timing was done randomly by test center, the centers rather than individuals were used as units for analysis. Independent analyses were conducted within each subgroup of interest (as identified by sex by race by "years out" categories). The subgroup's mean on the experimental test was the dependent variable in the analysis. In order to increase the precision of the estimated effect of timing for each group, a covariate (the mean GRE score corresponding to that of the experimental test) was used. The slopes of the dependent variables on the covariate were allowed to differ in the experimental and control timings, and an average effect of timing was estimated. This analysis thus produced an estimated effect and its standard error in each subgroup, and *t*-tests were used to test the significance of the differences between racial groups, sex groups, and "years out" categories.

Centers with fewer than five individuals were excluded from the data analysis for several reasons: (1) centers with small numbers of examinees would get a disproportional large weight in the estimation of effects; (2) most small centers would not contribute to the calculation of *t*-tests since at least one individual in each group being compared must be in the center in order for the center to contribute to the significance test; and (3) testing experience may be somewhat different in a small center. Eighteen centers were eliminated from the 20-minute verbal analyses, 18 from the 20-minute quantitative analyses, 9 from the 30-minute verbal analyses, and 8 from the 30-minute quantitative analyses.

Item Analysis. Item analysis refers to a set of analyses of examinee responses to questions on a test. An item analysis, performed when a test is first administered, determines whether the questions and the test as a whole are working as anticipated. This includes a calculation of standard measures of speededness frequently used to evaluate the appropriateness of time limits for standardized tests. Swineford (1956) established a set of criteria defining a test as unspeeded in three ways: (1) if virtually all candidates complete 75 percent of the items, (2) if at least 80 percent of the candidates reach the last item, and (3) if the variance of the items not reached does not exceed 25% of the total-score variance. The statistics to evaluate speededness, the mean and standard deviation

of the formula score, the mean of the number of omitted items, the mean and standard deviation of not-reached items, and the Kuder-Richardson formula 20 reliability estimates adapted for formula scoring were obtained separately for Black males, Black females, White males, and White females.

Any examinees who contributed to center estimates of effects were eligible for inclusion in the item analyses. All Black males and females were included, but in each item analysis for the White group a sample of every third examinee was used.

Results

Regression Analysis for Estimating Effects of Timing

The primary focus of this study was to determine whether increasing the time allotments for the GRE verbal and quantitative sections would differentially affect the scores of subgroups of interest. Summaries of the effects of increasing the time on the experimental tests, after controlling for the comparable operational score, are presented in the following sections. (Appendices A and B show summary statistics by individual rather than by center for readers who would like to compare the summary tables.) Since the assignment of timing was done randomly by center, the analysis was at the center level. For each center, means were calculated for the covariate (operational score on verbal or quantitative sections) and for the experimental test for the groups of subjects under analysis (e.g., female, white, 0-4 years out). The center was then used as the unit in the regression analysis from which effects were calculated. The effects for combined groups were calculated by the same technique, where the center mean is based on combined groups of subjects in the center (e.g., for female analyses, all females, regardless of ethnic group or years since the bachelor's degree, were used in calculating the center means). When the treatment effects are to be compared in groups A and B (e.g., males and females), each center is summarized by two values: the difference in covariate means between the A and B groups in the center and the difference in dependent variable means between the A and B groups in the center. These differences across test centers are the input to the regression analysis and consequently are used to test the interaction between treatment and group membership. This procedure properly allows for the data structure with the test center as a unit, since only centers with members of the groups compared are included in the significance test (i.e., the effects by cells are not assumed to be independent). The effects estimated with test center as a unit are appropriate to individuals. The analysis using test center as a unit merely estimates the effects and their standard errors in cognizance of the clustering (as with cluster sampling) of indivudals by centers.

Verbal. Summary statistics including number of centers, mean center experimental score, and standard deviations of center scores on the operational verbal and experimental verbal tests are presented in Table 1. The estimated average effect and standard error for increasing the time from 20 to 30 minutes on the experimental verbal test were obtained for each category of the Sex and Race and Years Out design and are presented in Table 2. A large number of t-tests were performed to identify significant differences in effects between cells. Of the 261 t-tests, 13 would be expected to be significant by chance alone. Only 7 significant differences were found. Table 3 presents a summary of these results. No significant differences in treatment effects were obtained within groups categorized by Sex or Race (Black, White) or number of years since baccalaureate degree (0-4, 5-9, 10-14, 15-19, 20-24). Groups categorized by Sex and Race also showed no significant differences when the appropriate comparisons of effects were made. Comparisons of groups categorized by Sex and Years (45 t-tests) resulted in only three significant differences in effect. Women 5-9 years out gained significantly more than women 0-4 years out and men 5-9 years out when given more time to complete the test. Women 20-24 years out gained significantly more than men 10-14 years out. One significant difference in effect was found when comparing groups categorized by race and years out. Given the longer timing, Whites 15-19 years out gained significantly more than Blacks 0-4 years out.

Groups categorized by Sex and Race and Years Out resulted in several significant differences in effect when t-tests were applied to the numerous possible comparisons. White women 15-19 years out increased scores more by increased timing than Black women 0-4 years out. Black women 20-24 years out gained significantly more than Black women 10-14 years out when given more time. White men 20-24 years out gained more by increased timing than the White men 10-14 years out. Of the 190 possible individual cell comparisons, 154 comparisons could be made (the number of centers was too small for the other tests.)

Quantitative. Table 4 contains the summary statistics for operational and experimental quantitative tests used in the regression analysis. The estimated average effect and standard error for increasing testing time on the experimental quantitative test is summarized in Table 5. Significant differences in effects between cells identified by a series of t-tests are summarized for the experimental quantitative test in Table 6. Of the 257 t-tests calculated, 13 would be significant by chance alone. Fifteen significant differences were actually identified. Significant differences in effects did not occur for groups categorized by Sex (1 t-test) or Race (1 t-test), or Years Out (10 t-tests). Groups categorized by Sex and Race also produced no significant differences when comparisons of effects were made. Comparisons of groups categorized by Sex and Years resulted in two significant differences in effect. Men 10-14 years out gained significantly more than women

TABLE 1
Regression Analysis Summary Statistics for Operational and Experimental Verbal Tests with Center as the Unit of Analysis

Years Out	Experimental Test Time		BLACK MALE		BLACK FEMALE		WHITE MALE		WHITE FEMALE		RACE AND SEX COMBINED	
			OPERATIONAL (95 questions)	EXPERIMENTAL (26 questions)	OPERATIONAL (95 questions)	EXPERIMENTAL (26 questions)	OPERATIONAL (95 questions)	EXPERIMENTAL (26 questions)	OPERATIONAL (95 questions)	EXPERIMENTAL (26 questions)	OPERATIONAL (95 questions)	EXPERIMENTAL (26 questions)
0-4	20-minute	N	103	103	122	122	246	246	240	240	250	250
		Mean	30.99	6.41	30.73	5.80	48.26	12.39	49.06	11.57	46.61	11.37
		SD	14.43	4.64	13.19	3.79	7.58	2.61	7.04	2.49	7.51	2.52
	30-minute	N	98	98	120	120	245	245	241	241	253	253
		Mean	30.73	7.73	29.82	6.17	47.93	13.24	50.31	12.65	46.92	12.35
		SD	12.91	4.37	11.07	3.82	7.63	2.48	7.58	2.85	8.31	2.72
5-9	20-minute	N	22	22	30	30	164	164	159	159	204	204
		Mean	25.79	5.52	32.24	5.76	51.86	13.85	53.25	12.99	49.62	12.66
		SD	13.80	3.91	15.84	5.09	14.09	4.15	14.13	4.47	13.25	3.99
	30-minute	N	22	22	31	31	155	155	146	146	204	204
		Mean	34.78	8.55	32.94	8.01	53.02	14.85	53.31	14.38	51.16	14.04
		SD	19.66	5.67	17.59	5.59	12.82	4.38	13.38	4.53	12.79	4.24
10-14	20-minute	N	5	5	13	13	90	90	95	95	140	140
		Mean	21.65	5.30	24.23	4.84	52.25	14.05	53.57	13.53	50.43	13.09
		SD	13.64	5.25	14.58	5.04	17.06	5.32	15.11	4.95	16.44	5.12
	30-minute	N	5	5	14	14	73	73	88	88	124	124
		Mean	39.75	10.85	29.53	7.04	53.26	15.09	56.96	15.49	53.04	14.57
		SD	13.15	5.11	15.59	5.57	16.83	4.91	16.24	5.56	16.24	5.33
15-19	20-minute	N	1	1	10	10	43	43	54	54	87	87
		Mean	35.13	10.50	32.45	7.88	48.01	13.13	52.46	12.98	49.55	12.86
		SD	—	—	17.37	6.30	18.49	5.85	16.41	5.69	17.32	5.67
	30-minute	N	6	6	12	12	33	33	58	58	86	86
		Mean	27.54	7.54	21.96	5.44	47.95	14.97	52.13	14.51	47.29	13.47
		SD	19.99	5.74	14.43	4.28	15.94	4.13	16.30	5.34	16.51	5.35
20-24	20-minute	N	6	6	7	7	22	22	32	32	58	58
		Mean	30.67	7.17	30.68	6.46	46.73	11.78	50.83	12.68	45.89	11.29
		SD	11.04	5.97	18.36	6.71	20.43	6.66	19.28	5.78	19.45	6.25
	30-minute	N	1	1	7	7	10	10	31	31	46	46
		Mean	28.75	5.25	29.11	7.77	38.35	10.73	51.80	15.10	45.26	12.93
		SD	—	—	13.78	3.61	17.83	4.21	19.50	5.96	18.77	5.55

Table 1
(Continued)

			BLACK		WHITE		RACE AND YEARS OUT COMBINED	
			OPERATIONAL (95 questions)	EXPERIMENTAL (26 questions)	OPERATIONAL (95 questions)	EXPERIMENTAL (26 questions)	OPERATIONAL (95 questions)	EXPERIMENTAL (26 questions)
MALE	20-minute	N	114	114	246	246	250	250
		Mean	30.07	6.19	48.52	12.53	46.87	11.99
		SD	13.88	4.51	7.58	2.52	8.43	2.84
	30-minute	N	109	109	247	247	253	253
		Mean	31.34	7.89	48.34	13.42	46.86	12.93
		SD	13.27	4.35	7.52	2.41	8.69	2.79
FEMALE	20-minute	N	128	128	244	244	248	248
		Mean	31.45	5.94	49.60	11.72	47.44	11.03
		SD	12.76	3.84	7.20	2.48	8.61	2.89
	30-minute	N	130	130	246	246	250	250
		Mean	29.70	6.28	50.66	12.89	48.18	12.15
		SD	10.92	3.78	7.70	2.73	9.24	3.02
SEX AND YEARS OUT COMBINED	20-minute	N	151	151	246	246	250	250
		Mean	31.04	6.06	48.84	12.14	46.95	11.51
		SD	12.05	3.82	6.33	2.06	7.64	2.50
	30-minute	N	154	154	249	249	253	253
		Mean	30.91	7.12	49.29	13.13	47.34	12.53
		SD	11.19	3.82	6.55	2.17	8.08	2.61

Table 2

Table of Effects for Increasing the Time from 20 to 30 Minutes on the Experimental Verbal Test

		BLACK		WHITE		
		MEN	WOMEN	MEN	WOMEN	RACE & SEX COMBINED
0-4	N	201	242	491	481	503
	effect	+1.39	+.60	+.95	+.69	+.88
	SE	±.36	±.29	±.10	±.12	±.08
5-9	N	44	61	319	305	408
	effect	+.99	+2.07	+.68	+1.37	+.95
	SE	±.93	±.80	±.24	±.29	±.20
10-14	N	10	27	163	183	264
	effect	+.12	+.69	+.78	+.99	+.74
	SE	±2.58	±1.16	±.42	±.41	±.30
15-19	N	7	22	76	112	173
	effect	-1.10	+.21	+1.84	+1.62	+1.24
	SE	±3.68	±1.51	±.75	±.53	±.43
20-24	N	7	14	32	63	104
	effect	-1.30	+1.76	+1.06	+2.16	+1.82
	SE	±5.79	±1.02	±1.35	±.69	±.58

		BLACK	WHITE	RACE AND YEARS OUT COMBINED
MEN	N	223	493	503
	effect	+1.36	+.94	+.94
	SE	±.34	±.10	±.09
WOMEN	N	258	490	498
	effect	+.79	+.84	+.89
	SE	±.28	±.11	±.11
SEX AND YEARS OUT COMBINED	N	305	495	503
	effect	+1.09	+.85	+.90
	SE	±.26	±.07	±.07

Table 3

Summary of Significant Differences for the Experimental Verbal Tests

	Significant Differences	Number of Testable Differences*	Number of Possible Significant Differences
Sex (collapsing over Race & Years)	0	1	1
Race (collapsing over Sex & Years)	0	1	1
Years (collapsing over Sex & Race)	0	10	10
Sex and Race (collapsing over Years)	0	6	6
Sex and Years (collapsing over Race)	3	45	45
Race and Years (collapsing over Sex)	1	44	45
Sex and Race and Years	3	154	190
TOTAL	7	261	298

*Because of insufficient degrees of freedom (N=centers) all possible testable differences were not calculable.

Table 4
Regression Analysis Summary Statistics for Operational and
Experimental Quantitative Tests with Center as the Unit of Analysis

Years Out	Experimental Test Time		BLACK MALE OPERATIONAL (55 questions)	BLACK MALE EXPERIMENTAL (14 questions)	BLACK FEMALE OPERATIONAL (55 questions)	BLACK FEMALE EXPERIMENTAL (14 questions)	WHITE MALE OPERATIONAL (55 questions)	WHITE MALE EXPERIMENTAL (14 questions)	WHITE FEMALE OPERATIONAL (55 questions)	WHITE FEMALE EXPERIMENTAL (14 questions)
0-4	20-minute	N	105	105	116	116	251	251	248	248
		Mean	18.50	4.08	14.46	3.32	33.86	8.13	26.99	6.15
		SD	8.74	2.38	7.53	2.09	5.09	1.58	4.84	1.37
	30-minute	N	102	102	115	115	249	249	247	247
		Mean	19.63	4.85	14.66	3.79	33.65	8.75	27.34	6.96
		SD	8.65	2.83	7.08	2.08	5.39	1.73	4.89	1.51
5-9	20-minute	N	20	20	41	41	178	178	167	167
		Mean	14.68	2.67	13.88	3.36	32.41	8.18	26.44	6.31
		SD	7.18	2.16	8.68	2.56	8.61	2.60	7.22	2.53
	30-minute	N	20	20	34	34	175	175	143	143
		Mean	15.39	4.34	12.45	3.15	31.64	8.62	26.15	7.10
		SD	8.73	2.86	7.11	1.95	9.61	2.86	7.78	2.65
10-14	20-minute	N	9	9	20	20	68	68	100	100
		Mean	13.81	3.53	13.26	2.63	29.37	7.28	24.27	6.25
		SD	6.17	2.93	7.12	2.38	10.07	2.84	9.17	3.15
	30-minute	N	16	16	15	15	78	78	86	86
		Mean	14.33	5.08	11.49	3.59	31.09	8.56	24.82	6.80
		SD	8.44	3.43	3.48	2.40	10.37	3.15	8.17	2.95
15-19	20-minute	N	5	5	8	8	38	38	55	55
		Mean	16.20	3.95	14.28	1.77	28.92	7.09	20.40	5.42
		SD	6.75	2.42	13.05	1.90	9.26	3.27	8.17	2.64
	30-minute	N	5	5	9	9	31	31	48	48
		Mean	18.90	5.80	8.64	2.39	28.26	7.98	22.97	7.06
		SD	12.79	4.26	7.86	2.63	10.70	3.96	7.62	2.49
20-24	20-minute	N	3	3	5	5	11	11	33	33
		Mean	17.33	5.17	11.18	1.40	33.22	8.88	18.93	4.49
		SD	10.18	3.76	6.90	1.90	12.38	3.52	8.88	2.98
	30-minute	N	1	1	4	4	17	17	25	25
		Mean	10.00	3.75	6.75	3.75	27.94	7.63	21.65	6.44
		SD	—	--	3.03	2.21	10.54	4.16	8.32	3.48

Table 4
(Continued)

			BLACK		WHITE		RACE AND YEARS OUT COMBINED	
			OPERATIONAL (55 questions)	EXPERIMENTAL (14 questions)	OPERATIONAL (55 questions)	EXPERIMENTAL (14 questions)	OPERATIONAL (55 questions)	EXPERIMENTAL (14 questions)
MALE	20-minute	N	116	116	251	251	255	255
		Mean	17.94	4.06	33.67	8.15	32.40	7.81
		SD	8.14	2.30	4.69	1.45	5.36	1.54
	30-minute	N	110	110	251	251	255	255
		Mean	19.40	4.96	33.33	8.73	32.23	8.45
		SD	8.47	2.82	5.25	1.65	6.01	1.81
FEMALE	20-minute	N	131	131	250	250	255	255
		Mean	13.96	3.08	26.54	6.09	25.17	5.77
		SD	7.20	2.07	4.65	1.37	5.38	1.46
	30-minute	N	126	126	250	250	254	254
		Mean	13.98	3.69	26.96	6.96	25.62	6.62
		SD	6.62	2.07	4.96	1.47	5.82	1.60
SEX AND YEARS OUT COMBINED	20-minute	N	166	166	251	251	255	255
		Mean	16.00	3.54	30.29	7.18	28.91	6.83
		SD	7.72	2.02	4.18	1.22	5.17	1.41
	30-minute	N	155	155	252	252	255	255
		Mean	16.48	4.38	30.27	7.86	28.98	7.54
		SD	7.44	2.36	4.47	1.40	5.45	1.54

Table 5

Table of Effects for Increasing the Time from 20 to 30 Minutes on the Experimental Quantitative Test

		BLACK		WHITE		
		MEN	WOMEN	MEN	WOMEN	SEX & RACE COMBINED
0-4	N	207	231	500	495	510
	effect	+.50	+.44	+.68	+.73	+.67
	SE	±.22	±.21	±.06	±.08	±.05
5-9	N	40	75	353	310	425
	effect	+1.54	+.02	+.63	+.87	+.84
	SE	±.62	±.46	±.16	±.18	±.12
10-14	N	25	35	146	186	280
	effect	+1.37	+1.37	+.87	+.41	+.71
	SE	± .80	± .70	±.30	±.29	±.21
15-19	N	10	17	69	103	155
	effect	+1.00	+1.40	+1.08	+1.08	+1.10
	SE	±.86	±1.01	±.50	±.39	±.30
20-24	N	4	9	28	58	88
	effect	+1.24	+2.72	+.35	+1.29	+1.14
	SE	±1.27	±1.58	±.67	±.68	±.51

		BLACK	WHITE	RACE & YEARS OUT COMBINED
MEN	N	226	502	510
	effect	+.55	+.68	+.69
	SE	±.21	±.06	±.05
WOMEN	N	257	500	509
	effect	+.61	+.78	+.74
	SE	±.20	±.07	±.07
SEX AND YEARS OUT COMBINED	N	321	503	510
	effect	+.74	+.69	+.69
	SE	±.16	±.05	±.04

Table 6

Summary of Significant Differences for the Experimental Quantitative Tests

	Significant Differences	Number of* Testable Differences	Number of Possible Testable Differences
Sex (collapsing over Race & Years)	0	1	1
Race (collapsing over Sex & Years)	0	1	1
Years (collapsing over Sex & Race)	0	10	10
Sex and Race (collapsing over Years)	0	6	6
Sex and Years (collapsing over Race)	2	45	45
Race and Years (collapsing over Sex)	6	40	45
Sex and Race and Years	7	154	190
TOTAL	15	257	298

*Because of insufficient degrees of freedom (N=centers) all possible testable differences were not calculable.

10-14 years out when given more time to complete the test. Women 15-19 years out gained more than women 5-9 years out when given the longer timing.

Groups categorized by Race and Years resulted in six significant differences when the appropriate comparisons of effects were made. Of the 45 possible comparisons, only 40 could be tested. Blacks 10-14 years out increased scores more when given more time than Blacks 0-4 years out, Blacks 5-9 years out or Whites 20-24 years out. Blacks 20-24 years out gained more when given more time than Whites 0-4 years out or Whites 20-24 years out. Blacks 15-19 years out increased scores more than Blacks 0-4 years out when given more time.

Cells categorized by Sex and Race and Years Out resulted in seven significant differences in effect when compared. Black men 10-14 years out gained significantly more by the longer timing than Black men 0-4 or White men 10-14 years out. Black women 10-14 years out did significantly better than Black men 0-4 years out, Black women 5-9 years out, or White men 15-19 years out when given more time. Black men 20-24 years out gained significantly more than White women 0-4 years out when given more time to complete the test. White men 10-14 years out gained more by increased timing than the White women 10-14 years out. Sufficient numbers of centers were available to test only 155 of the 190 possible comparisons.

Summary of Effects. Only 7 significant differences were found for the verbal experimental test and 15 significant differences for the quantitative experimental test (see Tables 3 and 6). Of the approximately 260 *t*-tests calculated for each experimental test, about 13 would be significant by chance using an alpha level of .05. The use of such a large number of *t*-tests to analyze a single data base results in a high probability of rejecting a true null hypothesis (Type I error). Tables 2 and 5 might suggest that there is an overall trend for increasing effects as the number of years since receipt of baccelaurete increases. We therefore did *t*-tests for the linear trend, assuming that the estimates for each years-out category are independent. The t-values for verbal and quantitative are 1.20 and .74 respectively. Thus, this is not a significant trend and the few significant results can probably be attributed to random occurences.

Item Analysis

Verbal. Sample sizes, means, standard deviations, speededness statistics, and reliabilities are presented for experimental verbal test item analyses groups in Table 7. The White examinees have higher mean scores than the Black examinees. Within race and sex categories, mean scores were higher on the 30-minute than on the 20-minute versions. Results of the three speededness measures vary slightly, but generally the 20-minute version is somewhat more speeded than the 30-minute version for all groups. The 20-minute

Table 7
Item Analysis Summary Data for the Verbal Experimental Tests
(26 questions)

RACE/ SEX	TIME	N	MEAN	S.D.	Percentage of Candidates Completing 75 Percent of the Items	Number of Items Completed by 80 Percent of the Candidates	Ratio of Not Reached Variance to Total Score Variance	Percentage of Candidates Completing the Last Item	OMITS MEAN	OMITS S.D.	NOT REACHED MEAN	NOT REACHED S.D.	Reliability
BLACKS													
MALE	20-minute	295	5.95	4.75	76	17	1.07	22	2.84	3.32	4.55	4.92	.74
	30-minute	265	7.34	4.99	95	22	0.31	41	2.14	2.51	2.05	2.80	.73
FEMALE	20-minute	570	5.12	4.68	84	20	0.75	29	2.71	3.07	3.52	4.04	.73
	30-minute	515	5.87	4.74	95	23	0.30	40	2.37	2.65	1.89	2.60	.71
WHITE													
MALE	20-minute	1,490	13.15	5.43	95	23	0.21	31	1.95	2.12	2.01	2.51	.79
	30-minute	1,440	14.27	5.38	100	23	0.08	38	1.46	1.78	1.34	1.51	.79
FEMALE	20-minute	1,460	12.11	5.54	95	23	0.22	27	1.98	2.34	2.15	2.60	.79
	30-minute	1,365	13.06	5.62	99	24	0.07	39	1.34	1.81	1.27	1.47	.79

verbal experimental test appears most speeded for the Black males. Table 8 presents comparable statistics for the operational verbal sections. Section 1 appears to be more speeded than section 2. It is difficult to estimate whether the 20-minute experimental section is comparable to the two separately timed sections. Although it appears roughly comparable for the White groups, the 20-minute sections appear much more speeded for the Black groups than would be indicated by the operational section item analyses. The differences may be due to unreliability in the speededness estimates.

Quantitative. Sample sizes, means, standard deviations, speededness statistics, and reliabilities for the experimental quantitative test item analysis groups are presented in Table 9. Mean scores are highest for White males, followed by White females, Black males, and Black females. Within ethnic-sex categories, mean scores on the 20-minute versions are lower than mean scores on the 30-minute version. Generally the test appears more speeded for Black than for White examinees. Table 10 presents comparable statistics for the operational quantitative section of the test. On the basis of the information in Tables 9 and 10, it is difficult to estimate whether the 20-minute experimental section is comparable in speededness to the operational section, since the speededness measures are not controlled for initial ability and do have some measurement error themselves.

Summary of Verbal and Quantitative Item Analysis. A summary of the results of the study using the three speededness criteria for verbal and quantitative experimental tests is presented below.

Table 11
Summary of Speededness Criteria by Race and Sex

	Verbal						Quantitative					
	20-minutes			30-minutes			20-minutes			30-minutes		
Sex/Race	Criteria			Criteria			Criteria			Criteria		
	1	2	3	1	2	3	1	2	3	1	2	3
Male Black	S	S	S	S/U	S	S	S/U	S	S	U	U	U
Female Black	S	S	S	S/U	S	S	S/U	S	S	U	S	S
Male White	S/U	S	U	U	S	U	U	S	U	U	U	U
Female White	S/U	S	U	U	S	U	S/U	S	S	U	S	U

S = Speeded U = Unspeeded S/U = Borderline

The results shown in Table 11 should be interpreted with caution. Interpretation should be in the context of sample sizes and the mean scores presented in Tables 7 and 9. The groups compared have large sample-size differences in some instances; thus speededness measures

Table 8
Item Analysis Summary Data for the Verbal Operational Sections

RACE/ SEX	Experimental Section TIME	N	MEAN	S.D.	Percentage of Candidates Completing 75 Percent of the Items	Number of Items Completed by 80 Percent of the Candidates	Ratio of Not Reached Variance to Total Score Variance	Percentage of Candidates Completing the Last Item	OMITS MEAN	OMITS SD	NOT REACHED MEAN	NOT REACHED S.D.	Reliability
SECTION 1 - OPERATIONAL (55 questions)													
BLACKS													
MALE	20-minute	295	12.76	7.92	91.5	47	0.68	44.4	13.46	10.07	4.33	6.54	.82
	30-minute	265	13.49	7.98	90.6	48	0.60	49.1	12.47	10.02	3.94	6.19	.81
FEMALE	20-minute	570	12.58	7.90	91.1	49	0.67	55.3	11.94	9.88	3.62	6.46	.81
	30-minute	515	12.55	7.64	89.3	49	0.85	56.1	11.91	9.60	3.92	7.04	.80
WHITE													
MALE	20-minute	1490	24.03	9.56	97.0	52	0.19	65.6	11.33	8.06	1.91	4.18	.87
	30-minute	1440	24.42	9.76	97.2	52	0.17	68.1	11.39	8.20	1.77	4.02	.88
FEMALE	20-minute	1460	24.81	9.62	97.1	52	0.19	68.7	10.74	7.60	1.78	4.16	.88
	30-minute	1365	25.00	9.79	96.0	52	0.22	67.3	10.36	7.73	2.05	4.56	.88
SECTION 2 - OPERATIONAL (40 questions)													
BLACKS													
MALE	20-minute	295	15.52	8.52	90.8	33	0.35	58.6	2.78	4.56	3.22	5.07	.85
	30-minute	265	16.11	8.62	95.5	34	0.21	60.0	2.36	3.63	2.45	3.95	.85
FEMALE	20-minute	570	15.38	8.36	93.0	34	0.23	62.5	2.41	3.76	2.48	4.01	.84
	30-minute	515	14.84	8.23	91.8	33	0.26	56.9	2.66	3.88	2.87	4.18	.84
WHITE													
MALE	20-minute	1490	26.10	6.86	98.7	40	0.12	83.0	1.28	2.47	0.86	2.37	.80
	30-minute	1440	26.47	6.87	99.2	40	0.09	84.2	1.38	2.71	0.72	2.07	.80
FEMALE	20-minute	1460	26.01	6.87	99.5	40	0.08	82.6	1.46	2.74	0.75	1.98	.80
	30-minute	1365	26.21	6.92	99.0	40	0.11	82.6	1.39	2.69	0.84	2.27	.80

Table 9
Item Analysis Summary Data for the Quantitative Experimental Tests
(14 questions)

RACE/ SEX	TIME	N	MEAN	S.D.	Percentage of Candidates Completing 75 Percent of the Items	Number of Items Completed by 80 Percent of the Candidates	Ratio of Not Reached Variance to Total Score Variance	Percentage of Candidates Completing the Last Item	OMITS MEAN	OMITS S.D.	NOT REACHED MEAN	NOT REACHED S.D.	Reliability
BLACKS													
MALE	20-minute	310	3.98	2.95	92	11	0.47	61	1.59	1.71	1.33	2.03	.64
	30-minute	305	4.77	3.44	97	13	0.16	80	1.41	1.62	0.58	1.39	.72
FEMALE	20-minute	590	3.13	2.50	92	11	0.67	65	1.69	1.70	1.22	2.05	.50
	30-minute	535	3.42	2.60	96	12	0.45	71	1.66	1.79	0.86	1.79	.52
WHITE													
MALE	20-minute	1475	8.49	3.27	96	12	0.23	70	1.32	1.45	0.89	1.58	.75
	30-minute	1440	9.40	3.41	99	14	0.10	88	0.92	1.28	0.33	1.09	.77
FEMALE	20-minute	1445	6.52	3.06	93	11	0.41	59	1.80	1.66	1.35	1.95	.70
	30-minute	1380	7.25	3.29	97	13	0.20	76	1.43	1.49	0.68	1.47	.73

Table 10
Item Analysis Summary Data for the Quantitative Operational Tests
(55 questions)

RACE/ SEX	Experimental Section TIME	N	MEAN	S.D.	Percentage of Candidates Completing 75 Percent of the Items	Number of Items Completed by 80 Percent of the Candidates	Ratio of Not Reached Variance to Total Score Variance	Percentage of Candidates Completing the Last Item	OMITS MEAN	OMITS S.D.	NOT REACHED MEAN	NOT REACHED S.D.	Reliability
BLACKS													
MALE	20-minute	310	17.60	10.44	94.5	51	0.27	43.9	8.65	7.91	2.94	5.40	.89
	30-minute	305	18.47	10.43	95.1	49	0.26	40.3	7.70	7.28	3.30	5.36	.89
FEMALE	20-minute	590	13.33	8.59	94.4	50	0.43	44.1	9.40	7.62	3.11	5.65	.84
	30-minute	535	12.63	8.35	94.2	51	0.44	43.0	9.88	8.16	3.03	5.57	.83
WHITE													
MALE	20-minute	1475	34.38	10.79	99.0	54	0.08	58.2	4.47	5.35	1.42	3.06	.91
	30-minute	1440	35.21	10.91	99.1	54	0.09	64.2	4.19	5.37	1.32	3.33	.91
FEMALE	20-minute	1445	27.50	9.91	97.5	53	0.14	48.6	7.47	6.46	1.90	3.68	.88
	30-minute	1380	27.51	9.75	96.9	53	0.17	48.9	7.21	6.16	2.02	3.99	.88

are not consistently reliable across groups. In addition, it would be expected that a test is more speeded for a lower-scoring group. There may be a relationship between the mean scores on the experimental verbal and quantitative tests and the degree to which the members of the groups did not complete the test. It should be noted that for the experimental verbal and quantitative tests the items were ordered as to difficulty--from easy at the beginning to most difficult at the end. The test was also scored with a correction for guessing. Thus, the last items in the test may have been intentionally omitted by examinees, thus exaggerating the speededness estimates.

Summary and Recommendations

The major purpose of the study was to determine if increasing the time available per question has a differential effect on the scores (after controlling for initial ability) of examinee subgroups as defined by sex, race, and the number of years since receipt of the baccalaureate degree. By using the GRE score on the corresponding operational section to control for differences in group ability, the differential impact of timing on scores was investigated. The overall average effect of increasing the time by 10 minutes on the verbal experimental test is less than one point on a 26-question test. The overall average effect of increasing the time on the 14 questions of the quantitative experimental test is even less, about two-thirds of a point. These would correspond to about 3.3 and 2.7 formula-score points on the respective operational sections, or 20 to 30 points on the reported scores. In neither case were average effects significantly different for males versus females, for Blacks versus Whites, or for various levels of years since the receipt of the baccalaureate degree.

Thus, although a larger proportion of examinees complete the experimental tests when given additional time, this extra time does not differentially help any of the groups studied. There could be several reasons for this. First, extra time allows examinees who have completed the test to review their answers. This idea is supported in a study by Mueller and Wasser (1971) in which they concluded that examinees who change answers on tests gain more points than they lose. Second, lower scoring examinees find the test more difficult and tend not to complete the test. Given more time to answer questions, lower scoring examinees would be expected to answer fewer of the additional questions, especially since the more difficult test items appear at the end.

Results of this study indicate that test-taking time is not a likely "biasing agent" on the GRE test scores for subgroups defined by sex, race, or the number of years since their receipt of the baccalaureate degree. Increased test-taking time would increase the scores of all groups to about the same degree.

References

Evans, F. R., & Reilly, R. R. The LSAT Speededness Study--Revisited. Report #LSAC-72-3. In Law School Admissions Council, Reports of LSAC Sponsored Research: Volume II, 1970-74, Princeton, N.J.: Law School Admission Council, 1976. (a)

Evans, F. R., & Reilly, R. R. A study of speededness as a source of test bias. Journal of Educational Measurement, 1972, 9(2), 123-131. (b)

Evans, F. R., & Reilly, R. R. A study of test speededness as a potential source of bias in the quantitative score of the Admissions Test for Graduate Study in Business. Research in Higher Education, 1973, 1, 173-183.

Mueller, D. J., & Wasser, V. Implications of changing answers on objective test items. Journal of Educational Measurement, 1977, 14, 9-14.

Swineford, F. Technical manual for users of test analyses (SR 56-42). Princeton, N.J.: Educational Testing Service, 1956.

Appendix A
Individual Summary Statistics for Operational and Experimental Verbal Tests

Years Out	Experimental Test Time		BLACK				WHITE			
			MALE		FEMALE		MALE		FEMALE	
			OPERATIONAL (95 questions)	EXPERIMENTAL (26 questions)	OPERATIONAL (95 questions)	EXPERIMENTAL (26 questions)	OPERATIONAL (95 questions)	EXPERIMENTAL (26 questions)	OPERATIONAL (95 questions)	EXPERIMENTAL (26 questions)
0-4	20-minute	N	254	254	506	506	4060	4060	3839	3839
		Mean	28.38	5.82	27.53	4.89	50.01	12.96	50.18	11.79
		SD	15.62	4.76	14.89	4.49	14.96	5.35	14.85	5.46
	30-minute	N	230	230	446	446	4022	4022	3628	3628
		Mean	28.85	7.07	26.50	5.44	50.03	13.88	50.54	12.78
		SD	14.75	4.88	14.13	4.42	14.90	5.30	14.79	5.46
5-9	20-minute	N	29	29	43	43	435	435	472	472
		Mean	25.19	5.33	30.25	5.51	52.35	13.86	53.40	13.18
		SD	14.19	4.20	16.36	5.43	17.08	5.65	16.50	5.35
	30-minute	N	29	29	41	41	429	429	446	446
		Mean	31.69	7.77	34.03	8.29	53.31	15.01	53.02	14.31
		SD	19.52	5.50	19.77	6.46	16.47	5.44	16.63	5.54
10-14	20-minute	N	5	5	16	16	118	118	157	157
		Mean	21.65	5.30	22.98	4.48	52.04	13.97	54.00	13.61
		SD	13.64	5.25	13.51	4.75	17.30	5.65	16.43	5.63
	30-minute	N	5	5	18	18	92	92	149	149
		Mean	39.75	10.85	28.33	6.67	52.72	15.14	57.42	15.40
		SD	13.15	5.11	17.27	5.95	18.48	5.44	17.52	5.90
15-19	20-minute	N	2	2	12	12	48	48	67	67
		Mean	35.12	10.50	30.42	7.12	48.01	13.11	51.41	12.84
		SD	15.73	2.47	16.86	6.08	18.18	6.06	17.81	6.04
	30-minute	N	6	6	12	12	40	40	84	84
		Mean	27.54	7.54	21.96	5.44	48.96	15.22	53.17	14.89
		SD	19.99	5.74	14.43	4.28	17.76	4.63	14.72	5.33
20-24	20-minute	N	6	6	7	7	23	23	35	35
		Mean	30.67	7.17	30.68	6.46	46.09	11.76	51.69	12.90
		SD	11.04	5.97	18.36	6.71	21.02	6.51	20.03	5.71
	30-minute	N	1	1	8	8	10	10	36	36
		Mean	28.75	5.25	28.56	7.38	38.35	10.72	53.21	15.35
		SD	—	—	13.58	4.02	17.83	4.21	19.52	5.74

Appendix A
(Continued)

			BLACK		WHITE		RACE AND YEARS OUT COMBINED	
			OPERATIONAL (95 questions)	EXPERIMENTAL (26 questions)	OPERATIONAL (95 questions)	EXPERIMENTAL (26 questions)	OPERATIONAL (95 questions)	EXPERIMENTAL (26 questions)
MALE	20-minute	N	296	296	4,684	4,684	4,980	4,980
		Mean	28.04	5.82	50.24	13.07	48.92	12.64
		SD	15.36	4.72	15.32	5.41	16.19	5.63
	30-minute	N	271	271	4,593	4,593	4,864	4,864
		Mean	29.33	7.21	50.35	14.02	49.18	13.64
		SD	15.39	4.96	15.20	5.32	15.96	5.52
FEMALE	20-minute	N	584	584	4,570	4,570	5,154	5,154
		Mean	27.70	4.99	50.67	12.02	48.07	11.22
		SD	15.04	4.64	15.21	5.49	16.85	5.84
	30-minute	N	525	525	4,343	4,343	4,868	4,868
		Mean	27.08	5.73	51.11	13.09	48.51	12.30
		SD	14.85	4.71	15.25	5.53	16.94	5.90
SEX AND YEARS OUT COMBINED	20-minute	N	880	880	9,254	9,254	10,134	10,134
		Mean	27.81	5.27	50.45	12.55	48.48	11.92
		SD	15.14	4.68	15.27	5.47	16.53	5.78
	30-minute	N	796	796	8,936	8,936	9,732	9,732
		Mean	27.85	6.24	50.72	13.57	48.85	12.97
		SD	15.06	4.85	15.23	5.44	16.46	5.76

Appendix B
Individual Summary Statistics for Operational and Experimental Quantitative Tests

Years Out	Experimental Test Time		BLACK				WHITE			
			MALE		FEMALE		MALE		FEMALE	
			OPERATIONAL (55 questions)	EXPERIMENTAL (14 questions)	OPERATIONAL (55 questions)	EXPERIMENTAL (14 questions)	OPERATIONAL (55 questions)	EXPERIMENTAL (14 questions)	OPERATIONAL (55 questions)	EXPERIMENTAL (14 questions)
0-4	20-minute	N	276	276	498	498	4036	4036	3853	3853
		Mean	17.85	3.94	13.19	3.04	34.81	8.47	27.73	6.44
		SD	10.77	2.99	8.58	2.48	10.57	3.25	9.93	2.99
	30-minute	N	262	262	451	451	3982	3982	3651	3651
		Mean	19.32	4.71	12.70	3.33	35.21	9.27	27.77	7.07
		SD	10.62	3.43	8.63	2.61	10.73	3.39	9.95	3.26
5-9	20-minute	N	23	23	61	61	447	447	444	444
		Mean	14.71	2.57	13.25	3.20	32.28	8.11	26.81	6.58
		SD	7.68	2.16	8.43	2.56	10.29	3.18	9.15	3.00
	30-minute	N	26	26	55	55	466	466	447	447
		Mean	13.56	4.01	12.15	3.31	32.44	8.89	25.85	7.06
		SD	8.73	3.09	7.25	2.25	11.21	3.47	9.81	3.33
10-14	20-minute	N	9	9	23	23	87	87	160	160
		Mean	13.81	3.53	13.83	2.74	29.32	7.35	25.26	6.46
		SD	6.17	2.93	7.30	2.42	11.01	3.19	10.57	3.43
	30-minute	N	18	18	22	22	121	121	163	163
		Mean	13.24	4.65	10.41	3.27	31.24	8.70	25.34	6.93
		SD	8.70	3.60	4.25	2.38	10.99	3.41	9.35	3.36
15-19	20-minute	N	5	5	9	9	45	45	74	74
		Mean	16.20	3.95	13.97	1.61	28.78	7.16	20.71	5.52
		SD	6.75	2.42	12.37	1.96	9.51	3.40	8.62	2.99
	30-minute	N	5	5	9	9	38	38	64	64
		Mean	18.90	5.80	8.64	2.39	28.51	8.20	22.84	7.00
		SD	12.79	4.26	7.86	2.63	11.65	4.11	8.21	2.83
20-24	20-minute	N	3	3	6	6	13	13	38	38
		Mean	17.33	5.17	10.75	1.38	33.71	9.19	19.27	4.68
		SD	10.18	3.76	6.78	1.71	13.60	3.92	8.63	3.06
	30-minute	N	1	1	4	4	20	20	29	29
		Mean	10.00	3.75	6.75	3.75	28.16	7.64	21.27	6.34
		SD	--	--	3.03	2.21	10.75	3.95	8.70	3.40

Appendix B
(Continued)

			BLACK		WHITE		RACE AND YEARS OUT COMBINED	
			OPERATIONAL (55 questions)	EXPERIMENTAL (14 questions)	OPERATIONAL (55 questions)	EXPERIMENTAL (14 questions)	OPERATIONAL (55 questions)	EXPERIMENTAL (14 questions)
MALE	20-minute	N	316	316	4,628	4,628	4,944	4,944
		Mean	17.47	3.84	34.40	8.40	33.32	8.11
		SD	10.42	2.95	10.61	3.25	11.38	3.42
	30-minute	N	312	312	4,627	4,627	4,939	4,939
		Mean	18.45	4.67	34.74	9.21	33.72	8.92
		SD	10.56	3.42	10.87	3.41	11.55	3.58
FEMALE	20-minute	N	597	597	4,569	4,569	5,166	5,166
		Mean	13.21	3.01	27.37	6.43	25.73	6.03
		SD	8.55	2.48	9.93	3.01	10.77	3.15
	30-minute	N	541	541	4,354	4,354	4,895	4,895
		Mean	12.44	3.31	27.37	7.06	25.72	6.65
		SD	8.34	2.56	9.93	3.26	10.83	3.40
SEX AND YEARS OUT COMBINED	20-minute	N	913	913	9,197	9,197	10,110	10,110
		Mean	14.68	3.30	30.91	7.42	29.44	7.05
		SD	9.46	2.68	10.86	3.29	11.71	3.45
	30-minute	N	853	853	8,981	8,981	9,834	9.834
		Mean	14.64	3.80	31.17	8.17	29.74	7.79
		SD	9.65	2.97	11.06	3.51	11.89	3.68

Appendix C
Verbal
Comparisons Not Testable (Sex and Race and Years)

SEX	RACE	YEARS OUT	SEX	RACE	YEARS OUT
Male	Black	5-9	Male	Black	10-14
Male	Black	5-9	Male	Black	15-19
Male	Black	5-9	Male	Black	20-24
Male	Black	5-9	Female	Black	20-24
Female	Black	5-9	Male	Black	10-14
Female	Black	5-9	Male	Black	15-19
Female	Black	5-9	Male	Black	20-24
Male	Black	10-14	Female	Black	10-14
Male	Black	10-14	Male	Black	15-19
Male	Black	10-14	Female	Black	15-19
Male	Black	10-14	Male	White	15-19
Male	Black	10-14	Male	Black	20-24
Male	Black	10-14	Female	Black	20-24
Male	Black	10-14	Male	White	20-24
Male	Black	10-14	Female	White	20-24
Female	Black	10-14	Male	Black	15-19
Female	Black	10-14	Male	Black	20-24
Female	Black	10-14	Male	White	20-24
Male	White	10-14	Male	Black	20-24
Male	Black	15-19	Female	Black	15-19
Male	Black	15-19	Male	White	15-19
Male	Black	15-19	Female	White	15-19
Male	Black	15-19	Male	Black	20-24
Male	Black	15-19	Female	Black	20-24
Male	Black	15-19	Male	White	20-24
Male	Black	15-19	Female	White	20-24
Female	Black	15-19	Male	White	15-19
Female	Black	15-19	Male	Black	20-24
Female	Black	15-19	Female	Black	20-24
Male	White	15-19	Male	Black	20-24
Female	White	15-19	Male	Black	20-24
Male	Black	20-24	Female	Black	20-24
Male	Black	20-24	Male	White	20-24
Male	Black	20-24	Female	White	20-24
Female	Black	20-24	Male	White	20-24
Female	Black	20-24	Female	White	20-24

Appendix D
Verbal Marginal
Comparisons Not Testable

RACE	SEX	YEARS OUT	RACE	SEX	YEARS OUT
Black		15-19	Black		20-24

Appendix E
Quantitative
Comparisons Not Testable (Sex and Race and Years)

SEX	RACE	YEARS OUT	SEX	RACE	YEARS OUT
Male	Black	0-4	Male	Black	20-24
Female	Black	0-4	Male	Black	20-24
Male	Black	5-9	Male	Black	15-19
Male	Black	5-9	Female	Black	20-24
Male	Black	5-9	Male	Black	20-24
Female	Black	5-9	Male	Black	20-24
Female	Black	5-9	Female	Black	20-24
Male	Black	10-14	Male	Black	15-19
Male	Black	10-14	Male	Black	20-24
Male	Black	10-14	Female	Black	20-24
Male	Black	10-14	Male	White	20-24
Female	Black	10-14	Female	Black	15-19
Female	Black	10-14	Male	Black	20-24
Female	Black	10-14	Female	Black	20-24
Male	White	10-14	Male	Black	15-19
Male	White	10-14	Male	Black	20-24
Female	White	10-14	Male	Black	20-24
Male	Black	15-19	Male	White	15-19
Male	Black	15-19	Female	White	15-19
Male	Black	15-19	Male	Black	20-24
Male	Black	15-19	Female	Black	20-24
Male	Black	15-19	Male	White	20-24
Male	Black	15-19	Female	White	20-24
Female	Black	15-19	Male	Black	20-24
Female	Black	15-19	Female	Black	20-24
Female	Black	15-19	Male	White	20-24
Female	Black	15-19	Female	White	20-24
Male	White	15-19	Male	Black	20-24
Male	White	15-19	Female	Black	20-24
Female	White	15-19	Male	Black	20-24
Female	White	15-19	Female	Black	20-24
Male	Black	20-24	Female	Black	20-24
Male	Black	20-24	Male	White	20-24
Male	Black	20-24	Female	White	20-24
Female	Black	20-24	Male	White	20-24
Female	Black	20-24	Female	White	20-24

Appendix F
Quantitative Marginal
Comparisons Not Testable

RACE	SEX	YEARS OUT	RACE	SEX	YEARS OUT
Black		5-9	Black		20-24
Black		10-14	Black		20-24
Black		15-19	Black		20-24
White		15-19	Black		20-24
White		15-19	Black		20-24

GRE BOARD RESEARCH
REPORTS OF A TECHNICAL NATURE

Boldt, R. R. Comparison of a Bayesian and a Least Squares Method of Educational Prediction. GRE Board Professional Report GREB No. 70-3P, June 1975.

Campbell, J. T. and Belcher, L. H. Word Associations of Students at Predominantly White and Predominantly Black Colleges. GRE Board Professional Report GREB No. 71-6P, December 1975.

Carlson, A. B.; Reilly, R. R.; Mahoney, M. H.; and Casserly, P. L. The Development and Pilot Testing of Criterion Rating Scales. GRE Board Professional Report GREB No. 73-1P, October 1976.

Carlson, A. B.; Evans, F.R.; and Kuykendall, N. M. The Feasibility of Common Criterion Validity Studies of the GRE. GRE Board Professional Report GREB No. 71-1P, July 1974.

Donlon, T. F.; Reilly, R. R.; and McKee, J. D. Development of a Test of Global vs. Articulated Thinking: The Figure Location Test. GRE Board Professional Report GREB No. 74-9P, June 1978.

Echternacht, G. Alternate Methods of Equating GRE Advanced Tests. GRE Board Professional Report GREB No. 69-2P, June 1974.

Echternacht, G. A Comparison of Various Item Option Weighting Schemes/A Note on the Variances of Empirically Derived Option Scoring Weights. GRE Board Professional Report GREB No. 71-17P, February 1975.

Echternacht, G. A Quick Method for Determining Test Bias. GRE Board Professional Report GREB No. 70-8P, July 1974.

Evans, F. R. The GRE-Q Coaching/Instruction Study. GRE Board Professional Report GREB No. 71-5aP, September 1977.

Fredericksen, N. and Ward, W. C. Development of Measures for the Study of Creativity. GRE Board Professional Report GREB No. 72-2P, June 1975.

McPeek, M.; Altman, R. A.; Wallmark, M.; and Wingersky, B. C. An Investigation of the Feasibility of Obtaining Additional Subscores on the GRE Advanced Psychology Test. GRE Board Professional Report GREB No. 74-4P, April 1976.

Powers, D. E.; Swinton, S.; Thayer, D.; and Yates, A. A Factor Analytic Investigation of Seven Experimental Analytical Item Types. GRE Board Professional Report GREB No. 77-1P, June 1978.

Powers, D. E.; Swinton, S. S.; and Carlson, A. B. A Factor Analytic Study of the GRE Aptitude Test. GRE Board Professional Report GREB No. 75-11P, September 1977.

Reilly, R. R. and Jackson, R. Effects of Empirical Option Weighting on Reliability and Validity of the GRE. GRE Board Professional Report GREB No. 71-9P, July 1974.

Reilly, R. R. Factors in Graduate Student Performance. GRE Board Professional Report GREB No. 71-2P, July 1974.

Rock, D. A. The Identification of Population Moderators and Their Effect on the Prediction of Doctorate Attainment. GRE Board Professional Report GREB No. 69-6bP, February 1975.

Rock, D. A. The "Test Chooser": A Different Approach to a Prediction Weighting Scheme. GRE Board Professional Report GREB No. 70-2P, November 1974.

Sharon, A. T. Test of English as a Foreign Language as a Moderator of Graduate Record Examinations Scores in the Prediction of Foreign Students' Grades in Graduate School. GRE Board Professional Report GREB No. 70-1P, June 1974.

Ward, W. C.; Frederiksen, N.; and Carlson, S. B. Construct Validity of Free-Response and Machine-Scorable Versions of a Test of Scientific Thinking. GRE Board Professional Report GREB No. 74-8P, November 1978.

Ward, W. C. and Frederiksen, N. A Study of the Predictive Validity of the Tests of Scientific Thinking. GRE Board Professional Report GREB No. 74-6P, October 1977.