GRE RESEARCH

Effects of Applying Different Time Limits to a Proposed GRE Writing Test

Donald E. Powers
Mary E. Fowles

ETS Educational Testing Service

Effects of Applying Different Time Limits

to a Proposed GRE Writing Test

Donald E. Powers

Mary E. Fowles

GRE Board Report No. 93-26cR

November 1997

Educational Testing Service, Princeton, NJ 08541

****

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

****

Acknowledgments

Abstract

To determine the role of time limits on both test performance and test validity, we asked approximately 300 prospective graduate students to each write two essays -- one in a 40-minute time period and the other in 60 minutes. Analyses revealed that, on average, test performance was significantly better when examinees were given 60 minutes instead of 40. However, there was no interaction between test-taking style (fast vs. slow) and time limits. That is, examinees who described themselves as slow writers/test takers did not benefit any more (or any less) from generous time limits than did their quicker counterparts. In addition, there was no detectable effect of different time limits on the meaning of essay scores, as suggested by their relationship to several nontest indicators of writing ability.

Effects of Applying Different Time Limits to a Proposed GRE Writing Test

Psychometricians have customarily distinguished between so-called power tests and speed tests. Power tests contain questions of varying difficulty and, in their purest form, afford all test takers ample time to consider and respond to every test question. In contrast, speed tests typically present only very easy questions that, if reached, can be answered correctly by virtually all test takers.

Although a power/speed dichotomy is useful theoretically, the distinction is seldom neatly observed in practice. For test administrations to be feasible, some time restrictions must usually be set, even for power tests. Therefore, while most standardized tests are intended primarily to reflect test takers' intellectual "power" (rather than the rate at which they work), virtually all such tests involve some element of speed, however minor.

In addition to purely practical reasons, a primary motivation for enforcing strict time limits is to ensure that all examinees face the same, standardized task, and thus that (at least in this sense) the test is fair for every test taker. Ironically, test makers' loyalty to stringent time allocations is often a vexing aspect of standardized testing -- both for critics of testing and for test takers themselves. For instance, in its report to Congress on educational testing, the Office of Technology Assessment (1992) reiterated a familiar lament -- that time constraints produce a situation that "... rewards the mentally quick and penalizes the ruminative" (p. 209). Also, from our experience with standardized testing, we know that test takers frequently complain about the "pressures of time," "rigid time constraints," and the lack of opportunity to "work at your own pace to show what you know."

Time limits pose still another kind of problem for direct measures of writing ability. Enforcing time limits may, some feel, diminish a writing test's "authenticity." A frequent assertion is that impromptu writing performed in a rigidly prescribed amount of time is not a very good representation of what test users value most. (An alternative view, however, once expressed by a graduate dean, is that often the ability to fashion impromptu responses, for example, in composing electronic mail messages or in taking timed course-related examinations, is _exactly_ what is needed.)

A source of anxiety for test takers and an issue for test critics, time limits are also a worry for test makers -- as a potential source of unfairness both to specific individuals and to particular subclassifications of test takers. The concern is that some individuals, or subgroups, may be disadvantaged because of their characteristic test-taking styles (e.g., slow and deliberate) or their particular cultural values (e.g., a low premium on quickness or a high priority on reflectivity). (See, for example, Llabre, 1991.) As a graduate faculty member once told us, typical writing tests may penalize certain students -- e.g., the "diligent researcher," who, while possibly unable to "write wonderfully on the spot," could perform quite well when given sufficient time.

Related Research

To guard against the possibility of unfairness, many major testing programs have conducted speededness studies to investigate whether additional time (or fewer questions) changes either the rank ordering of individual test takers or the magnitude of any between-group differences in their performance on multiple-choice tests. These studies have typically shown very high correlations between performance on more speeded and less speeded versions

of a test. Allowing more time normally results in better performance for everyone, but usually no particular subgroup benefits disproportionately from additional time. Studies of this kind have investigated subgroup performance according to age, gender, and ethnicity (e.g., Evans & Reilly, 1972, and Wild, Durso, & Rubin, 1982).

That test speededness seems to play little if any role in the differential performance of subgroups of examinees on multiple-choice tests is apparently due, at least partially, to the ordering of questions (from easiest to most difficult) that is typically employed. Because the final questions tend to be difficult for everyone, leaving them unanswered usually has little impact on examinees' test scores, at least in relation to those of other test takers. However, whereas this situation may hold true for most multiple-choice tests, it does not necessarily apply to tests (such as writing measures) that require a limited number of extended responses. Failing to provide the concluding paragraph of an expository essay or leaving insufficient time for editing a response could have greater consequences.

A number of studies have addressed the effects of time allotments for essay exams specifically. Some researchers (e.g., Biola, 1982; Hale, 1992; Younkin, 1986) have found that allowing more time results in higher scores. Other investigators (e.g., Livingston, 1987) have failed to detect any such effect. Small or nonsignificant effects have been noted even when comparisons have involved quite large differences in time allocations, for example, when essays were written in class versus when they were written at home over an extended period of time (several days or several weeks) (Caudery, 1990; Kroll, 1990).

Finally, one study bears mention because it included both the multiple-choice and the essay portion of an examination. Klein (1981) investigated the effects of extra time (55 vs. 90 minutes) for the California State Bar Examination and found a significant increase in average test scores for the essay portion, but little effect on the multiple-choice part of the examination.

With respect to subgroup performances, the research does not suggest that relaxing time limits significantly benefits any subgroup of examinees more than others. In most of these studies, however, the "groups of concern" have been defined by demographic characteristics (e.g., sex and ethnicity), not by any (possibly more relevant) psychological attributes. An exception is a study by Bridgeman (1980), who compared multiple-choice test scores of elementary students who were classified, on the basis of an independent measure, according to test-taking speed. Bridgeman found that test takers were not easily categorized and that students were not uniformly fast or slow across a variety of different tests.

Objectives

Because the GRE writing test is not intended to reflect the rate at which examinees work, the relevant classification variable is, we contend, slowness/deliberateness, a trait not confined to any particular demographic classification (though it may be more prominent in some groups than others). Therefore, the primary question that motivated the study reported here was, "Will slow, deliberate writers be disadvantaged by the time limits that must, for practical reasons, be observed when assessing writing in a large-scale testing program?"

The study was designed to meet the following objectives:

1. to determine how different time limits affect examinees' scores on a proposed GRE writing test
2. to investigate the extent to which slow, deliberate test takers may be at a disadvantage under the proposed time limits (and to describe who these people are likely to be)
3. to ascertain the relationship of test scores, under more and less stringent time conditions, to several nontest indicators of writing proficiency
4. to estimate any joint effects on test performance of (a) time limits and (b) disclosing essay topics before the test is administered

Method

Subjects

Study participants were 304 paid volunteers recruited from the pool of examinees who took the GRE General Test between January and May 1994. Subjects were selected from examinees who had tested at centers within a 50-60-mile radius of six ETS regional field offices (in Princeton, NJ; Washington, DC; Evanston, IL; Atlanta, GA; Phoenix, AZ; and Emeryville, CA) and from 10 GRE computer-based testing sites (in Arizona, Michigan, Minnesota, Nebraska, Texas, Virginia, and Wisconsin). Ethnic minority and non-native examinees were oversampled, and, in order to ensure sufficient heterogeneity with respect to writing ability, letters of invitation made a special plea to "students who do not consider their writing skills to be particularly strong." (The concern was that poor writers might not volunteer in numbers proportionate to their representation in the GRE test-taking population.)

Once identified from test-taker files, potential subjects were contacted with information about the requirements of the study and asked to indicate their interest in participating. Requirements consisted of taking a GRE writing test, completing a brief questionnaire, and submitting a typical example of writing they had done for a course. As part of a study of the effects of disclosing essay topics before the exam (see Powers & Fowles, in press), each participant also received copies of two topics that might be used on the test.

Instruments

Each participant wrote two full-length essays. Four topics were used in the study, as follows:

Topic A
"The absence of choice is very rare."
To what extent do you agree or disagree with the statement above? Support your position with reasons and examples drawn from areas such as your own reading and academic studies, your observation of people and events, or your personal experience.

Topic B
"What we call progress is a matter of exchanging one problem for another."
To what extent do you agree or disagree with the statement above? Support your
position with reasons and examples drawn from areas such as your own reading and
academic studies, your observation of people and events, or your personal experience.

Topic C
"Critical thinking and creative thinking are not opposite ends of a spectrum; they
are different facets of the same mind."
Using reasons and examples from your experience, observations, or reading, discuss the
extent to which you agree or disagree with the opinion expressed above. In your
discussion, be sure to make clear what you mean by "critical" and "creative" thinking.

Topic D
It is a common practice to attach "ism" to the end of a word in order to label a
particular movement, attitude, or trend. Such terms abound in almost every
sphere of life -- society, politics, the arts, philosophy, the sciences, and so on.
Choose -- or, if you prefer, invent -- one such "ism," define it, and explain why you
believe it is significant. Support your views with reasons and examples from your
experience, observation, and/or reading.

All participants completed a questionnaire to report, among other things, their
perceptions of the adequacy of time limits, an estimate of how quickly they are able to write,
and a judgment of how well they had performed on the writing tasks administered for our study.
The subjects were also asked to provide information about various aspects of their writing
ability, including (a) an overall assessment of their writing skills compared with those of other
students in their major field of study, (b) estimates of their success with various kinds of writing
(e.g., creative, persuasive, personal) and with various writing processes (e.g., organizing and
revising), and (c) undergraduate grades in courses requiring extensive writing and on their most
recent writing assignment. Some of these reports were solicited by administering an adaptation
of an inventory developed by Ward and Carlson (1984). We also asked study participants to
submit a course-related sample of writing (a term paper, book review, etc.) that they had
completed for an undergraduate assignment. Finally, several weeks after the test administration,
participants were recontacted and asked to complete a 12-item inventory of writing
accomplishments, that is, to indicate which of several writing accomplishments (e.g., publishing
an article or a letter to an editor) they had made. More detail about the nature of each of the
kinds of information described above is available in Powers, Fowles, and Boyles (in press).

Design

Each examinee wrote on two topics -- either topic A followed by topic B, or topic C
followed by topic D. (Each subject wrote on one topic that he or she had been sent before the
test and on one that he or she first encountered at the test administration.) Test forms (A/B or
C/D) were administered in a spiraled design so that approximately half of the subjects received
each form. Within each form, one of the essay topics was allotted 40 minutes and the other 60
minutes; half of the examinees wrote the 40-minute essay first, and half the 60-minute essay
first. Separate administrations were necessary for the two different timing conditions -- the 40-
minute/60-minute condition and the 60-minute/40-minute condition. Subjects were assigned

randomly in small groups to each timing condition at each testing site. After testing, each participant completed the questionnaire described above.

Scoring

The essays were scored holistically by a group of university faculty, all of whom were experienced essay scorers for large-scale writing assessment programs. These faculty were trained to apply the GRE scoring criteria to sample essays from this study before scoring began. A summary description of each score level appears below. (The complete scoring guide is available from the authors.) The 1-6 scale was anchored as follows:

6   Presents a thorough and insightful response to the question and demonstrates mastery of the elements of effective writing

5   Presents a well-developed response to the question and demonstrates a strong control of the elements of effective writing

4   Presents a clearly competent response to the question and demonstrates adequate control of the elements of writing

3   Presents a limited response to the question, either in content or control of the elements of writing

2   Presents a weak response to the question and demonstrates little control of the elements of writing

1   Is seriously deficient in writing skills

Each essay was scored independently by two trained readers. Discrepancies of two points or more were resolved by using a third reader. The correlations between two readers ranged from .77 to .80 for the four essay topics used in the study. Readers' agreement was virtually the same for essays administered with a 40-minute time limit ($r = .78$) as for those given with a 60-minute limit ($r = .77$).

The writing samples and personal statements submitted by participants were also scored on a 6-point holistic scale according to how well the writer communicated his or her ideas. Agreement on the quality of the writing contained in these samples was comparable to that achieved for the essays. (More detail about the scoring of these samples, as well as the problems encountered in doing so, is given in Powers, Fowles, & Boyles, in press.)

Designation of Subgroups

"Slowness/deliberateness," the examinee characteristic of interest, was operationalized as follows. Subjects were first classified as "slow," "average," or "fast" writers on the basis of responses to the following two questions:

1. Generally, how would you describe how quickly you are usually able to write reports, etc.? Compared with most other students, I probably write (much more slowly, somewhat more slowly, about the same, somewhat more quickly, much more quickly)

2. Please think about your past experiences with tests that have required you to write within a limited amount of time. How often have you felt pressured or frustrated because of difficulty finishing? (almost always, usually, sometimes, seldom, hardly ever)

Subjects were classified as slow writers (n=55) if they said that they (1) wrote reports either much more or somewhat more slowly than other students did and that they (2) either usually or almost always felt pressured or frustrated because of difficulty finishing. Those who said that they (1) wrote either somewhat more quickly or much more quickly and (2) seldom or hardly ever felt pressured or frustrated were considered to be fast (n=73). With the exception of 16 subjects who had inconsistent responses to the two questions (e.g., those who said they wrote much more quickly than others but who also said they were usually or almost always pressured or frustrated), all remaining subjects (n=141) were classified as average, neither fast nor slow. The crosstabulation of responses to these two questions is given in Table 1.

To ascertain the consistency with which subjects described themselves, we conducted a partial follow-up survey approximately four months later. Results were relatively consistent: for the 111 subjects who responded, the correlations between responses from the initial and follow-up surveys to the two questions were .74 and .66.

Table 1

Subjects' Reports of Feeling Pressured/Frustrated on Timed Writing Tests

Self-reported ability to write quickly (compared with fellow students)	Percentage feeling pressured/frustrated on timed writing tests				
	Almost Always	Usually	Sometimes	Seldom	Hardly Ever
Much slower (n=16)	44	44	0	6	6
Somewhat slower (n=80)	18	34	35	8	6
About the same (n=72)	3	14	56	21	7
Somewhat quicker (n=92)	4	7	39	37	13
Much quicker (n=41)	10	2	22	27	39

Validity of Subgroup Classification

To provide some objective validation of the categorization of test takers as slow, average, or fast, we obtained subjects' multiple-choice responses to the GRE General Test that they had taken previously. Using these data, we computed a score for the (approximately) first two thirds of each of the six separately timed sections of the test and also for the final third of each section. We made certain that the mix of different item types (e.g., reading comprehension, analogies, and antonyms) in the early portion of a test section was comparable to the mix in the later portion. Then, scores based on the later portion were regressed on scores from the early portion for each of the six test sections (two verbal, two quantitative, and two analytical). For each examinee, a residual (actual score minus predicted score on the later portion of the test) was computed for each section. Our rationale was that the later portions should reflect any test speededness more than the earlier portions. Thus, examinees who did better than expected on the later portion, as predicted from performance on the earlier portion, could be characterized

as being unaffected by time limits, relatively speaking, whereas those performing worse than expected could be regarded as being more susceptible to time limits.

The results of this analysis showed that the fast, average, and slow groups were ordered in the anticipated way for each of the six test sections: those classified as slow scored lower than predicted on the later portions, the average group scored about as expected, and the fast group scored higher than expected. However, because fast writers had higher GRE scores on average than did average writers, and slow writers had the lowest scores of any group, an alternative explanation is that the self-reports may have simply reflected ability level instead of, or in addition to, quickness.

In addition, we analyzed test takers' performance on individual test items (and on clusters of items) at various locations in the test, for example, items appearing early in the test versus ones appearing later. Fast test takers performed better than slow test takers on virtually every item (75 of 78) in the verbal sections of the test and on the vast majority (46 of 50) in the analytical sections. A majority of items (44 of 60) in the quantitative sections were also easier for fast than for slow test takers. There was, however, no clear tendency in any test section for multiple-choice items to be differentially more difficult for slow than for fast test takers.

The only possible exception was that, on average, reading comprehension questions were relatively more difficult for slow test takers. For example, on one verbal section of the test, slow test takers correctly answered 49% of reading comprehension questions and 56% of other verbal items, whereas fast test takers correctly answered 74% of reading comprehension items and 66% of other verbal items. It is commonly acknowledged that reading comprehension items are by far the most time-consuming of the verbal items on the GRE General Test. Test takers are often advised to answer the less time-consuming items first, and leave reading comprehension items for last, thus maximizing the return on their efforts. The relatively lower performance on reading comprehension by slow test takers may very well be a reflection of their running short on time at the end of the test, thus providing some evidence of the validity of our classifications.

Analyses

The main analyses were hierarchical regressions in which the dependent variable was essay score (summed over two readers). The independent variables, added sequentially in the following order, were
- time limit (40 or 60 minutes)
- disclosure (topic seen before testing or not)
- time limit by disclosure
- slowness/quickness (defined as three groups -- fast, average, and slow -- classified as described above, and coded as two dummy variables, S1 and S2)
- time limit by slowness/quickness

Our main interest was in determining the significance of the last interaction as an indication of the extent to which fast and slow examinees may benefit differentially from relaxing time limits, that is, allowing 50% more time (60 instead of 40 minutes).

We ran a second set of regression analyses to determine the extent to which different time limits influenced the relationship of essay scores to each of several nontest indicators of writing proficiency. The dependent variables were

(1) self-ratings of participants' performance on study tasks
(2) self-estimates of subjects' writing skill, compared with that of other students in their major field
(3) subjects' reports of success during college with each of a variety of writing activities and with several different kinds of writing assignments
(4) undergraduate grades, both in courses that required students to write and on the most recent writing assignment
(5) evaluations of course-related writing samples submitted by participants
(6) number of writing accomplishments reported

The independent variable in every case was the score on the essays students had written for the study. A product variable representing the interaction of time condition by essay performance was added last, and the significance of the increment in the multiple R squared due to this term was tested. Any significant effect was to be interpreted as follows: the association of essay performance with various indicators of writing proficiency depends on the amount of time allotted for essay writing.

All regression analyses were run as two separate replications. First, scores for the first essays written (i.e., A and C) were pooled for all subjects, and a concomitant variable was included to adjust for possible slight differences in the difficulty of essay topics A and C. Next, scores for the second essays written (topics B and D), were pooled and treated as a replication of the first analysis. This second analysis was a replication only in the sense that writing performance was based on new topics; the same subjects were used in both analyses.

Results

The Sample

Table 2 compares the study sample with all 1991-92 GRE General Test takers. As can be seen, the sample contains slightly higher proportions of women and arts/humanities majors than does the GRE population, and the GRE verbal and analytical scores of the sample are slightly higher, on average, than those of all GRE test takers. In most other respects, the study sample was comparable to the GRE test-taking population.

Reactions to Time Limits

Table 3 presents, by test-taker group, examinee reactions to time limits. Predictably, each group viewed the 60-minute limit as being more adequate than the 40-minute limit. In total, 75% of all subjects said that a 40-minute allocation was either just enough or more than enough time (versus not quite enough or not nearly enough), whereas 88% of the sample regarded 60 minutes to be either just enough or more than enough. More important though, there were significant differences in perceptions among the three test-taker groups: slower examinees judged each of the two limits to be less adequate than did faster examinees. A multivariate ANOVA (with reactions on a 1-4 scale as the dependent variable, time limit as a repeated measure factor, and test-taker group as a second factor) revealed the differences to be highly significant both between time limits, $F(1, 568) = 19.0$, $p < .001$, and among groups, $F(2, 568) = 20.5$, $p < .001$). The interaction of time limit and group was not significant, $F(2, 568) = 0.6$, n.s. These results are consistent, therefore, with our classification of test takers.

Table 2

Background Characteristics of Study Sample vs. GRE Test-Taking Population

Background Characteristic	Study Sample (n = 287)	1991-92 GRE Test-Taking Population (n > 277,000)
GRE General Test		
GRE verbal mean (SD)	514 (120)	486 (118)
GRE quantitative mean (SD)	544 (134)	555 (139)
GRE analytical mean (SD)	564 (121)	537 (129)
Female (%)	66	54
Ethnicity %		
American Indian	1	1
Asian American	7	3
Black	11	7
Hispanic*	3	4
White	77	83
Other	2	2
English Best Language (%)	89	N.A.
Degree Objective (% Doctorate)	39	41
Undergraduate Major Field (%)		
Business	6	4
Education	7	9
Arts/Humanities	24	16
Social Sciences	24	22
Engineering	7	14
Natural Sciences	21	27
Other	11	9
Undergraduate Grade Average (%)		
A	13	12
A-	29	26
B	34	37
B-	16	17
C	8	7
C- or lower	<1	<1

Note. GRE population data are based on unpublished statistics compiled from GRE General Test examinees' responses to the background information questionnaire that examinees complete when registering for the test. Percentages for the study sample are based on ns from 238 to 287.

* Includes Mexican American and other Hispanic.

Table 3

Percentages of Test Takers Perceiving Time Limits as Adequate
by Slow, Average, and Fast Test Takers

Time Limit	Slow	Average	Fast	p
40 Minutes	60	74	90	$p < .001$
60 Minutes	69	90	96	$p < .001$

Note. Entries are percentages of subjects who indicated they had just enough or more than enough time (vs. not quite enough or not nearly enough time). p is based on χ^2 test, using group by response (not nearly enough, not quite enough, just enough, more than enough).

Effects on Essay Scores

The results of the hierarchical regression analyses of essay scores are displayed in Table 4. For each replication (topics A and C and topics B and D), the effect of time allocation was highly significant, as was the effect of test-taker group (fast, average, slow). On average, essay scores were higher for essays written in 60 minutes than for those written in 40, and faster test takers performed better than slower ones. There was no interaction between time limit and disclosure condition or between time limit and test-taker group. The former result suggests that the better performance with 60 than 40 minutes did not depend on whether or not examinees had seen the topics before the exam. The latter suggests that none of the three test taker groups benefited significantly more than others when given additional time.

Descriptive statistics on the performance of each group under each time condition are presented in Table 5. The mean difference between fast and slow test takers (over both time conditions and all four essays) was 1.46 points on the 2-12 score scale, or an effect size (in standard deviation units) of .71. The mean difference in performance between essays written in 60 minutes and those written in 40 minutes was 0.9 points on the score scale, or an effect size of .43.

As suggested earlier, a major premise on which the study design was based is that such demographic characteristics as sex, age, and ethnic background are inappropriate surrogates for slowness/deliberateness in test-taking style, and that the real variable(s) of interest might best be considered more directly. Background characteristics are of interest, however, insofar as they may be correlated, however imperfectly, with test-taking style. Table 6 shows that each of the test-taking style groups is well represented by examinees from each demographic category. In fact, test-taking style was unrelated to sex, ethnicity, age, or undergraduate major. Subjects who said that English was not their best language, however, were represented disproportionately in the slower groups.

Table 4

Hierarchical Regressions for Time Limit Effects

Explanatory Variable Added	Cumulative R^2	Increase in R^2	F for Increase in R^2	df
	Topics A and C Pooled			
Time Limit (T)	.04	.04	10.2**	1, 279
Disclosure (D)	.04	.01	2.7	1, 278
D x T	.05	.01	2.3	1, 277
Slowness/Quickness[1] (S1, S2)	.12	.06	10.1***	2, 275
T x (S1, S2)	.12	.00	0.3	2, 273
	Topics B and D Pooled			
Time Limit (T)	.06	.06	16.3***	1, 279
Disclosure (D)	.06	.00	0.1	1, 278
D x T	.06	.01	1.9	1, 277
Slowness/Quickness[1] (S1, S2)	.12	.06	9.5***	2, 275
T x (S1, S2)	.13	.01	1.4	2, 273

Note. Topic was entered first as a covariate in each analysis.

[1] Subjects were classified as slow, average, or fast on the basis of responses to two survey questions. Slow was coded as -1, -1 for S1 and S2, average as 1, 0, and fast as 0, 1. Inconsistent responders ($n = 16$) were excluded from this analysis.

* $p < .05$, ** $p < .01$, *** $p < .001$

Table 5

Descriptive Statistics (Essay Scores) by Time Condition and Speed of Test Takers

Group		Time Limit	
		40 Minutes	60 Minutes
		Topics A and C	
Slow	M	6.6	7.7
	SD	2.3	2.0
	N	25	30
Average	M	7.1	8.0
	SD	2.2	1.9
	N	78	78
Fast	M	8.5	8.9
	SD	1.5	1.8
	N	33	40
		Topics B and D	
Slow	M	7.2	7.4
	SD	1.7	2.4
	N	30	25
Average	M	7.0	8.2
	SD	2.0	2.2
	N	78	78
Fast	M	8.2	9.2
	SD	1.5	1.4
	N	40	33

Table 6

Demographic Characteristics of Slow, Average, and Fast Test Takers

Characteristic		Slow	Average	Fast	p*
Female (%)		62	70	60	n.s.
Ethnicity (%)					
American Indian	(n=2)	2	1	0	
Black	(n=27)	10	13	8	
Mexican American	(n=4)	2	2	0	
Asian	(n=17)	10	8	3	n.s.
Other Hispanic	(n=2)	2	1	0	
White	(n=82)	73	73	86	
Other	(n=4)	0	2	3	
English Not Best Language (n=31) (%)		19	13	1	p<.01
Age (Mean)		30	28	28	n.s.
Undergraduate Major (%)					
Natural Sciences	(n=52)	23	23	14	
Engineering	(n=19)	6	8	8	
Social Sciences	(n=61)	23	26	22	n.s.
Arts/Humanities	(n=60)	25	24	23	
Education	(n=19)	8	6	11	
Business	(n=14)	6	3	11	
Other	(n=27)	8	11	12	

*p based on either χ^2 or F test.

The three groups did differ, however, on a host of other nondemographic variables, as revealed in Table 7, with faster examinees having better performances. For example, there were significant differences among the three groups with respect to GRE verbal and analytical scores, but not quantitative scores. All groups reported virtually identical overall average undergraduate grades, but differed with respect to grades in courses requiring writing and grade on the most recent writing assignment. Significant differences among groups were also noted with regard to (1) self-assessment of writing ability and (2) performance on the writing tasks completed for the study. Not shown in Table 7 are participants' assessments of their success in college with a variety of writing tasks and writing activities. On almost every task and activity, the three groups were ordered from slow to fast with respect to reported success. For instance, a total of 58%, 63%, and 90% of slow, average, and fast test takers, respectively, said they had been quite successful or extremely successful in developing an effective writing style. And 44%, 64%, and 93%, respectively, said they had been this successful on college examinations that involved long essays. Overall, most differences among the three groups suggested that our clustering (based on examinee self-reports of facility in writing and difficulty in completing timed tests) may have differentiated test takers as much on writing ability as on speed, possibly because the two constructs are not completely unrelated.

Effects on Correlations with Other Measures

Table 8 displays the correlations of several nontest indicators of writing proficiency with performances on essays administered under the two different time limits. To test for differences in these relations between the two time conditions, a series of regression analyses was run (for topics A and C together and for topics B and D together) in which each of the nontest indicators was, in turn, regressed on essay score and timing condition. Next, a product variable representing the interaction of essay score and timing condition was entered, and the significance of the contribution of this term to the multiple R squared was tested. A significant contribution would reveal that the relation between the indicator and essay performance depended on the amount of time that subjects were given to write their essays. A stronger relation under one timing condition than the other would suggest that writing skill was assessed more validly under one time limit than another. For none of the indicators, however, did the interaction variable contribute significantly to explaining variation in writing scores. Thus, our analysis reveals that an extra 20 minutes of writing time (60 minutes rather than 40) neither improved nor diminished the validity of the test, in terms of correlations of test scores with the various nontest indicators of writing skill considered here.

Discussion

To recapitulate the major findings of the study, time limits, even within a relatively limited range (40 vs. 60 minutes), do matter for the direct measure of writing that was investigated here. This 50% difference in time allotment had a detectable and practically significant effect on test performance: essays written in 60 minutes received moderately higher scores, on average, than did those written in 40 minutes. As noted earlier, some previous studies have noted similar effects from increasing time limits; others have not.

Table 7

Comparison of Slow, Average, and Fast Test Takers on Selected Characteristics

Characteristic	Slow	Average	Fast	p
Mean GRE General Test Performance				
Verbal	472	511	552	$p = .001$
Quantitative	535	537	564	n.s.
Analytical	512	566	598	$p < .001$
Reported Undergraduate Grades (Mean)				
Overall*	5.5	5.5	5.5	n.s.
In courses requiring writing*	5.6	5.8	6.4	$p < .001$
On most recent writing assignment*	6.2	6.3	6.7	$p = .001$
Self-Comparison of Writing Ability with That of Other Students				
% well above average	18	11	45	
% somewhat or well below average	15	6	0	$p < .001$**
Self-Rating of Writing Completed for This Study				
% well developed or thorough and insightful	25	29	59	
% seriously deficient, weak, or limited	45	28	4	$p < .001$**

*Scale is 1 = D or lower to 7 = A.

**Probabilities are for χ^2 test using all five response categories. Only selected response categories are displayed here.

Table 8

Correlations of Essay Scores with Several Indicators of Writing Proficiency

(by Timing Condition)

Indicator	40 Minutes	60 Minutes
Self-Assessments		
Overall self-rating of performance on study writing tasks	.37	.33
Comparison of writing with that of fellow students	.44	.39
Reported Success		
With various writing activities in college courses	.31	.27
With various kinds of writing in college courses	.36	.30
Independent Writing Sample		
Reader-graded (n=237)	.12	.19
Professor-graded (n=185)	.13	.18
Grades		
Overall average	.13	.13
Courses requiring writing	.30	.32
Most recent writing assignment	.33	.37
Writing accomplishments (n=142)	.31	.31

Note. All four topics have been pooled here. n=approximately 300 for each indicator unless otherwise specified. All correlations (except .12 and .13 for the 40-minute essay with the writing sample) are significant at the .05 level or beyond (two-tailed). None of the differences between correlations for the 40- and 60-minute essays was significant (p>.05).

The degree to which volunteer examinees in our study described themselves as able to write quickly and to perform without pressure on timed writing tests was strongly related to test performance, but unrelated to sex, age, ethnicity, or major field of study. Subjects who said that English was not their best language, however, were less likely than other examinees to describe themselves as being able to write quickly and to perform without pressure.

Although examinees' perceptions of themselves as slow or fast writers/test takers appeared to be relatively reliable, we were not able to provide any especially convincing validation of these perceptions. A plausible alternative interpretation of these perceptions (which we were unable to discount) is that perceptions may have reflected ability level as much

as they did the rate at which test takers actually work. Moreover, examinees may have been differentially able to recognize their slowness and/or differentially willing to admit it. Whether or not they correspond to reality, however, perceptions themselves are critical to examinees' acceptance of the writing test as a measure that is fair to all test takers.

There is a considerable body of research suggesting that general ability and speed, for example, in processing information, are in fact quite highly correlated (e.g., Vernon, 1987). Our data suggest that proficiency and quickness in writing are also quite highly related in our sample. Differences among the groups designated in this study represented not only differences in quickness, but very likely differences in writing proficiency. This assertion is supported by a number of study subjects' reports about their writing skills and achievements. In a sense, then, the study results may reveal the impact of different time limits on able and less able writers as much as they inform us about effects on slow and fast writers.

With respect to perceptions, faster test takers were more likely than slower examinees to judge time allocations (both 40- and 60-minute limits) as adequate. Differences among groups in perceptions of the adequacy of time limits remained relatively constant, however, regardless of which time limit was applied. That is, just as slower test takers considered 60-minute limits more acceptable than 40-minute limits, so too did faster test takers prefer the longer time limits (and to the same degree). (It is very likely, of course, that extremely generous limits might have been regarded as equally sufficient by fast and slow test takers.)

The most important finding from the study was that additional time was equally beneficial (and shorter limits equally disadvantageous) to slow, average, and fast test takers (or, more appropriately, to test takers who regarded themselves as slow or fast). That is, the relative performance of fast, average, and slow test takers did not change noticeably when more time was allowed. (Incidentally, these findings are consistent with anecdotal reports from test developers and faculty readers, who indicate that, when given extra time in which to write, less able writers often tend, simply, to continue to write poorly rather than improve what they have already written.)

Although, this study was somewhat limited in its scope. For instance, the investigation was constrained to a relatively restricted range of time limits. Thus, the findings will not assuage all critics, in particular those who have contended that

> even a substantially increased time allocation for completing a test task does not
> alter the fact that the students are being required to write under constraints that
> do not normally apply to the writing process (Read, 1991, pp. 87-88).

The findings do, however, suggest a certain independence (or "robustness") of the proposed GRE writing test with respect to relatively small, but practically meaningful, differences in time limits. Within the time limits studied here, this independence applies both to examinee perceptions of the stringency of time limits and to the examinees' actual test scores. Furthermore, the meaning of test scores, as defined by correlations with other nontest indicators of writing proficiency, was unrelated to time limit. In conclusion, based on the data collected here, scores on the writing measure were positively affected by allowing more time in which to write. However, neither the meaning of test scores nor the extent to which they distinguished among fast, average, and slow writers/test takers was related to moderate changes in time limits.

References

Biola, H. R. (1992). Time limits and topic assignments for essay tests. Research in the Teaching of English, 16, 97-98.

Bridgeman, B. (1980). Generality of a "fast" or "slow" test-taking style across a variety of cognitive tasks. Journal of Educational Measurement, 17, 211-217.

Caudery, T. (1990). The validity of timed essay tests in the assessment of writing skills. ELT Journal, 44, 122-131.

Congress of the United States, Office of Technology Assessment (1992). Testing in American schools: Asking the right questions (Summary). Washington, DC: Author.

Evans, F. R., & Reilly, R. R. (1972). A study of speededness as a source of test bias. Journal of Educational Measurement, 9, 123-31.

Hale, G. (1992). Effects of amount of time allowed on the Test of Written English (ETS Research Report-92-27). Princeton, NJ: Educational Testing Service.

Klein, S. P. (1981). The effect of time limits, item sequence, and question format on applicant performance on the California Bar Examination. San Francisco: Committee of Bar Examiners of the State Bar of California and the National Conference of Bar Examiners.

Kroll, B. (1990). What does time buy? ESL student performance on home versus class compositions. In B. Kroll (Ed.), Second language writing: Research insights for the classroom. Cambridge, England: Cambridge University Press.

Livingston, S. A. (1987, April). The effects of time limits on the quality of student-written essays. Paper presented at the meeting of the American Educational Research Association, Washington, DC.

Llabre, M. M. (1991). Time as a factor in the cognitive test performance of Latino college students (pp. 95-101). In C. D. Keller, J. R. Deneen, & R. J. Magallán (Eds.), Assessment and access: Hispanics in higher education. Albany: State University of New York Press.

Powers, D. E., & Fowles, M. E. (in press). Validating a new GRE writing test: Effects of disclosing essay topics. (GRE Report 93-26a). Princeton, NJ: Educational Testing Service.

Powers, D. E., Fowles, M. E., & Boyles, K. (in press). Validating a GRE writing test. (GRE Report 93-26b). Princeton, NJ: Educational Testing Service.

Read, J. (1991). The validity of writing test tasks. In S. Anivan (Ed.), Current developments in language testing. Singapore: SEAMEO Regional Language Centre.

Vernon, P. A. (Ed.) (1987). Speed of information processing and intelligence. Norwood, NJ: Ablex.

Younkin, W. F. (1986). Speededness as a source of test bias for non-native English speakers on the College Level Academic Skills Test (Doctoral dissertation, University of Miami). Dissertation Abstracts International, 47,/11-A, 4072.

Ward, W. C., & Carlson, S. B. (1984). A profile of preparation in English: Phase II (ETS Research Report No. 84-16 and College Board Report No. 84-2). New York: College Entrance Examination Board.

Wild, C. L, Durso, R., & Rubin, D. B. (1982). Effect of increased test-taking time on test scores by ethnic groups, years out of school, and sex. Journal of Educational Measurement, 19(2), 19-28.