GRE®

RESEARCH

Validating a Writing Test for Graduate Admissions

Donald E. Powers
Mary E. Fowles
Kelli Boyles

November 1996

GRE Board Professional Report No. 93-26bP
ETS Research Report 96-27

Educational Testing Service, Princeton, New Jersey

Validating a Writing Test for Graduate Admissions

Donald E. Powers

Mary E. Fowles

Kelli Boyles

GRE Board Report No. 93-26bP

November 1996

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 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.

****

The Graduate Record Examinations Board and Educational Testing Service are dedicated to the principle of equal opportunity, and their programs, services, and employment policies are guided by that principle.

Abstract

Proficiency in writing is important to success in many fields of graduate study. Measuring this construct, however, involves facing difficult challenges and criticisms from various perspectives. Among them is a concern that the writing skill that students demonstrate in an assessment situation may not relate fully enough to the writing competencies used in other contexts. This study was designed to reveal some of the facets of writing proficiency that different kinds of timed writing exercises might reflect. The approach taken was to examine how performance on these exercises correlates with performance on the following nontest indicators of writing skill: (a) self-assessments of writing skills, (b) self-reports of writing-related accomplishments and relevant undergraduate grades, and (c) independently graded samples of writing, such as course-related writing assignments and personal statements written in conjunction with application to graduate school. The writing exercises investigated in this study were of several kinds: composing a relatively traditional expository essay, writing somewhat briefer responses involving either explaining the meaning of a quotation or presenting ideas in a specified format, and revising a poorly written composition. These exercises were selected by university faculty members serving on a GRE Writing Committee.

A geographically diverse sample of 470 undergraduate students, all of whom had taken the GRE General Test (and who represented the GRE test-taking population reasonably well), were recruited to participate in the study. Subgroups of the study sample tried various combinations of the writing tasks.

The assorted nontest indicators of writing proficiency correlated with one another in differing degrees. None, however, appeared to be (statistically speaking) completely redundant with the others, suggesting that, in all likelihood, each reflected a somewhat different aspect of writing skill. The alternative writing exercises, in turn, exhibited distinctly different patterns of correlations with the individual indicators and with several composite indices based on logically determined subsets of indicators. By virtue of these distinct correlational patterns, the exercises appeared to have differential utility for a GRE writing test. A traditional issue-based expository essay appeared to be the most promising candidate. An exercise involving the revision of a composition also appeared to have potential. Two other experimental tasks, however, did not, from their relationships with the writing indicators, seem as defensible.

Limitations of the study are discussed -- principally the shortcomings of the nontest indicators as criteria for validating the assessment of writing proficiency. Despite the limitations, the results of the study are believed to have provided a reasonably good beginning to accumulating the kind of evidence required by professional standards for educational and psychological testing. Additional information will be required, however, about the validity of the inferences to be drawn and the decisions to be made about the examinees who will eventually take the GRE writing test.

Acknowledgments

The authors are grateful to the following people who contributed to the study reported here:

Phyllis Murphy, Patricia Bakr, Elizabeth Finotti, Dorice Hartman, Debera Hill-Walker, Susan Krupka, Rose Marie Meath, and Robert Wilmot from ETS Field Service Offices, who helped arrange for and carry out data collection

Robin Pergament and Lou Woodruff, who helped us recruit additional subjects and arrange for data collection

Kathy Carbery and Diane Thomas who matched files for us and retrieved GRE records

Elana Broch for consulting on project design

Seth Weiner for help in developing scoring methods

Faculty members John Collins, Marilyn Collins, Noreen Duncan, Abigail Harris, Ron Kostar, RoseAnn Morgan, and Marti Palmere for scoring the essays and the writing samples

Peggy Redman for preparing the data for analyses

Laura Jerry for analyzing the data

Ruth Yoder for carrying out numerous valuable administrative tasks

Hunter Breland, Brent Bridgeman, and Ann Willard for thoughtful reviews of an earlier draft

Members of both the GRE Research Committee and the GRE Writing Committee for their input to the study.

Validating a Writing Test for Graduate Admissions

The ability to communicate, that is, "to share one's ideas, knowledge, and insights with others" (both orally and in writing), (Enright & Gitomer, 1989, p. 10) is, according to eminent psychologists and distinguished scholars, a trait that is critical to successful performance in graduate school (Enright & Gitomer, 1989; Reilly, 1976), and undoubtedly in many professions as well. Communication skills are sometimes assessed quite informally, for example, during unstructured interviews or by casually skimming samples of applicants' writing. Often, however, more rigorous procedures are called for.

Strategies for testing writing skills have traditionally been of two general types: (1) indirect assessment, for which examinees typically answer multiple-choice questions about the conventions of writing, and (2) direct assessment, for which examinees are required to produce actual samples of writing as a demonstration of their skills. The merits of these approaches have been discussed in detail elsewhere (e.g., Breland & Gaynor, 1979; Cooper, 1984). Regardless of the particular strategy taken, however, measuring writing proficiency in any reasonably comprehensive manner poses a significant challenge for large-scale testing programs such as the Graduate Record Examinations (GRE). As cognitive psychologists have cautioned, "... the writing process is not a neat, additive sequence of stages" (Flower & Hayes, 1980, p. 44). It is not, therefore, amenable to easy capture, especially within the constraints of a typically short, one-time testing session.

Opinions differ widely about the feasibility of testing writing skills. Some (e.g., Wiggins, 1993) believe that, at least within very short time limits, the task is not possible at all:

> Whatever assessors are testing in a twenty-minute essay [such as found in some tests] ... most certainly is <u>not</u> the ability to write (p. 235).

Others, who have advocated the use of collections, or portfolios, of students' writing for both instruction and evaluation, have disputed the view that writing is a testable skill at all or that "...we can expect students or anyone else to write on demand" (Hout, 1994, p. 3).

The use of general "all-purpose" writing measures is also often grounds for criticism. Some faculty are often more interested in assessing discipline-specific skills than generic competencies. Moreover, the expected style of writing may differ among specialties according to the context and purpose for writing. Creative writing, for example, most assuredly invites the use of different skills than does a step-by-step demonstration of mathematical proofs. These observations were reinforced by the findings of a 1992 GRE survey of graduate deans and faculty, who often noted the discipline-specific nature of the writing required of graduate students.

Perhaps the best that can be hoped for as a result of giving a relatively brief standardized measure is a reasonably clear "snapshot" of an examinee's skills, rather than a full-length "video." When properly taken, however, even a simple photograph can often be remarkably revealing of its subject, provided that the photographer's aim is true and that extraneous influences do not unduly distort the image. In fact, faculty advisory committees and others involved in the development of writing assessments have found that many important aspects of writing skill can be captured with a test that is properly designed and administered.

Objectives

The major objective of the study reported here was to reveal some of the facets of writing ability that a carefully designed writing test may in fact reflect. Our approach was guided by a supposition that there is no single, compelling criterion against which to gauge the usefulness of scores from a measure of writing proficiency. This statement applies, of course, more generally to predicting success in graduate education, as Harnett and Willingham (1980) have noted in their discussion of the strengths and weaknesses of a variety of criteria of success. But, as Wainer (1993) has suggested, the lack of a single convincing indicator of the construct of interest (in this case, writing proficiency) does not necessarily signify that a composite of loosely correlated indicators will not be informative. The aim here was to explore the relationships of several standardized writing exercises to an assortment of nontest indicators of writing proficiency.

A secondary objective was to assess the comparative validity of a small set of writing tasks being considered by the GRE Writing Committee for inclusion in the new GRE writing test. Besides an expository essay, a variety of shorter, more experimental tasks was also of interest to the Committee.

Related Research

Informative reviews of validity studies of direct measures of writing have been provided by Breland (1983) and by Miller and Crocker (1990). Breland reviewed some two dozen studies, based mainly on high school and first-year college students, that reported correlations of writing assessments with concurrent measures or with measures taken some time later. The median correlation for concurrent measures, mostly high school or college grades, was .23. One study included in the Breland review (Huddleston, 1954) reported a correlation of .41 between instructors' ratings of high school students' writing ability and their performances on an essay test. Breland also found that correlations between essay test scores and scores on the same measure obtained 6-12 months later (reported for two studies) were in the .50s. Later grades, usually based only on courses in English, correlated in the .20s and low .30s.

Miller and Crocker (1990) cited studies showing correlations of .47 to .64 between direct measures of writing and the Test of Standard Written English (TSWE), a multiple-choice test of sentence structure and the conventions of standard written English. A study by Breland, Camp, Jones, Morris, and Rock (1987) -- characterized by Miller and Crocker as "the most comprehensive criterion-related validity study [of a direct measure of writing]" (p. 293) -- reported moderate correlations with instructors' ratings of students' writing ability (.41 to .57) and with course grades (.38 to .52) for students at several undergraduate institutions.

Method

Subjects

Study participants were 475 paid volunteers, recruited from examinees who took the GRE General Test at domestic test centers between January and May 1994. Selection was restricted to examinees who had tested at centers within a 50-60 mile radius of Educational Testing Service regional field offices in Atlanta, GA, Emeryville, CA, Evanston, IL, Phoenix,

AZ, Princeton, NJ, and Washington, DC, and at 10 computer-based testing centers in Arizona, Michigan, Minnesota, Nebraska, Texas, Virginia, and Wisconsin. Ethnic minority examinees and non-native speakers of English were oversampled, and, in order to enlist a sufficiently heterogeneous sample with respect to writing ability, letters of invitation made an explicit appeal to "students who do not consider their writing skills to be particularly strong." This request was designed to reduce the likelihood of a skewed sample that might result from the reluctance of poor writers to volunteer. Once identified, invitees were informed about the requirements of the study -- taking a GRE writing test, completing a brief questionnaire, and submitting a sample of writing they had done for a course and/or a personal statement they might have submitted to graduate schools -- and asked if they would like to participate.

Procedure

Approximately two to three weeks before testing, each participant received two essay topics along with a description of both the nature of the topics and the procedures that would be used to score the essays. A set of suggestions (like those anticipated to be included in test familiarization materials when the test becomes operational) on how to prepare for the test was also sent to each study participant.

Testing at each site was conducted in several small-group sessions consisting of approximately three to four examinees each. Volunteering subjects were assigned to groups at random. Approximately two-thirds of all participants were asked to write two issue-based expository essays (drawn from four topics included in the study) and to complete one of six shorter 20-minute writing exercises, which were alternated so that each exercise was given to every sixth examinee. The other third of participants were asked to attempt only the 20-minute exercises -- all six of them.

Examinees who wrote expository essays were given 40 minutes to write on one topic and 60 minutes to write on the other. Half of the subjects took the 40-minute essay first and half the 60-minute essay first. Topics were alternated so that approximately equal numbers of subjects wrote on each of four different topics in each timing condition. In addition, for each subject, one topic was one that had been seen before the test and one was previously unseen prior to testing. The effects of these different timing and pre-test disclosure conditions have been reported in detail elsewhere (Powers & Fowles, in press). In short, however, longer time limits yielded higher scores than did shorter ones, but pre-test disclosure had only a negligible effect.

Instruments

Expository Essays (40 or 60 minutes). Four essay topics were used in the study. Each topic asked students to explain their thoughts about a fairly complex issue, such as the following:

> "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.

Present Your Ideas in a Specified Format (20 minutes).

"In a successful relationship, both parties must be willing to compromise."

Write a unified paragraph in which you perform the following tasks:

-- Explain what you think the statement quoted above means.
-- Explain your own position on the subject.
-- Give an example to support your position.

Explain the Meaning of a Quotation (20 minutes).

"Talking is the worst possible way to get things done."

Briefly explain what you think the quotation above means and give an example to illustrate your interpretation.

Revise a Composition (20 minutes). Examinees were given a poorly written passage of about 350-400 words and asked to "revise and edit the passage so that it corresponds to the conventions of standard written English," treating it as a first draft of one of their own compositions. Examinees were instructed not to remove any ideas or to add any, but rather to "concentrate on finding weak places that need changing." (This exercise was not suggested by the GRE Writing Committee but rather was included at the request of the GRE Research Committee.)

Other Exercises. Three other kinds of exercises were also administered. However, because the GRE Writing Committee subsequently decided that these exercises were less promising than the others, they were dropped from further consideration, and therefore are not included in the analysis. These exercises involved (1) discussing the relationships between two statements or alternative views; (2) describing an object, event, process, etc., vividly and precisely, and (3) reflecting on the examinee's previously written essay about the meaning of a quotation (the examinee was asked, for example, to anticipate a criticism of his or her essay and to describe and refute it).

Questionnaire. After testing, all study participants completed a questionnaire (See Appendix A), providing, among other information:

1. perceptions of their prior success with different kinds of writing and with various writing tasks

2. information about undergraduate grades (overall average, average in courses requiring writing, and grade on the most recent assignment for which a written product was required) and

3. a self-evaluation of how the examinee's writing skills compare with those of peers in the same field of study.

Scoring

With the exception of the "Revise a Composition" task (for which only a subset was scored twice), each exercise was scored independently by two trained readers, with discrepancies of more than one point resolved by a third reader. The longer, expository "Issues" essays and one of the shorter exercises were scored on a 1-to-6 holistic score scale, ranging from 1= seriously deficient in writing skills to 6= presents a thorough and insightful response to the question and demonstrates mastery of the elements of effective writing. A second 20-minute exercise (reflecting on the "Meaning of a Quotation") was, after much deliberation, scored on a 3-point scale, as readers felt that only this coarser level of discrimination could be justified. The third 20-minute exercise ("Revise a Composition") was scored on a scale ranging from 1= deficient ("little or no improvement") to 6= highly effective revision ("major problems in logic and clarity effectively resolved").

Scoring the student-submitted writing samples presented a particularly interesting challenge. Study participants were given considerable latitude in the kind of writing sample they could submit. The request was for a recently written sample of 10 (or fewer) pages. Acceptable examples included essays, term papers, book reviews, lab reports, proposals, and case studies. The 10-page limit was frequently surpassed, however; samples ranged from highly specialized documents resembling computer programs to others that approached short autobiographies. Needless to say, the diversity of samples posed a significant challenge for scorers. Because the precise nature of the assignments for which the samples were written was not known explicitly, scoring was necessarily even more unorthodox. The development of separate scoring criteria for different types of papers -- by academic discipline, for example -- might have yielded more focused and more comparable scores for similar papers. However, this approach was not deemed practical for this study, given the extremely varied nature of the samples provided.

Consequently, all writing samples were scored on a 6-point holistic scale, according to how well the writer communicated his or her ideas to the university faculty trained as readers for this study. Essentially, readers were asked to evaluate the samples as if they had they been written for a "general academic community of readers." Readers were permitted to judge a paper as being unscorable if they regarded their own background knowledge as insufficient for understanding the paper's content. As a result, a few highly technical papers (mathematical proofs, lab reports, etc.) were left unscored. Besides the scores assigned by readers, grades awarded by professors were reported by students and used in subsequent analyses.

Personal statements were scored on the same 6-point scale used for essays and course-related writing samples. Readers were encouraged to assume the role of graduate admissions committee members, who might receive statements written by applicants from a variety of disciplines. Readers were directed to evaluate the statements for the quality of the writing exhibited, not for other indications of desirable personal traits that might be gleaned from the statements. The ways in which writers presented themselves (voice and tone) were regarded as relevant features to be considered.

Indicators of Writing Ability

As stated above, the general strategy was to relate performance on the prospective writing tasks to several different indicators, each of which might reflect a slightly different aspect of writing skills. Three broad categories were differentiated, as follows:

1. self-evaluations, that is, student perceptions of their own writing skills

2. self-reports of writing-related achievements and accomplishments

3. readers' assessments of samples of writing submitted by students

Information in the first category came from subjects' responses to several items in a questionnaire that was administered to all participating students. The questionnaire included an adaptation of an inventory developed by Ward and Carlson (1984), who asked secondary students how successful they had been with various processes of writing and with sundry kinds of writing assignments. Ward and Carlson found that such reports correlated in the .30s with a number of other measures, including the College Board's Test of Standard Written English, a measure of the knowledge of conventions of standard written English (.39).

Information in the second category was derived from responses to a 12-item inventory of writing accomplishments. This inventory was an abbreviated version of a longer, more comprehensive measure based on the work of Baird (1979) and extended by Lawrence Stricker. The particular adaptation used here involved only writing-related accomplishments, and respondents were not asked to provide documentation of their achievements, as was required by Baird.

The second category also included information about students' grades -- both in courses that involved writing and on the most recent assignment for which a report or similar exercise was required. Reports of overall undergraduate grade averages were also collected.

The third category included evaluations of writing samples (personal statements and course-related writing assignments) that, as mentioned above, were supplied by students and evaluated by trained readers for this study. Professor-assigned grades, as reported by students, were also examined as a basis for evaluation. A summary description of indicators within each category is given in Table 1.

Results

The Sample

Table 2 compares selected characteristics of the study sample with those of all GRE General Test takers (1991-92), the population of interest. Generally, the study sample appears to be reasonably representative of the GRE test-taking population. Differences are apparent with regard to the slightly higher proportion of women and arts/humanities majors in the study sample relative to the GRE population, and the somewhat higher average GRE verbal and analytical scores of the study sample.

Table 1

Summary Description of Independent Indicators of Writing Skills

Category	Description
Self-evaluation (Perceptions)	
Comparison with peers	"How do you think your writing compares with that of other students in your field?" (Responses on a 5-point scale from "well below average" to "well above average.")
Success with the processes of writing	Reported success in college courses with four different writing activities -- thinking about assignments, organizing, revising, and developing an effective writing style. (Responses on a 5-point scale ranging from "not at all successful" to "extremely successful.")
Success with various kinds of writing	Reported success in college courses with seven different kinds of writing -- personal, creative, persuasive, analytical/critical, descriptive, examinations, and applied. (Responses were on a 5-point scale ranging from "not at all successful" to "extremely successful.")
Evaluation of responses to study exercises	"How would you rate the writing you did for the tasks you were given [for this research study]?" (Responses were on a 6-point scale that corresponded to the holistic scale used to score essays.)
Self-reports of Achievements/Accomplishments	
Grade average in courses involving writing	Grade average in courses that required at least some writing.
Grade on most recent writing assignment	Grade on the most recent assignment for which a report, essay, etc., was required.
Documentable accomplishments in writing	The number of writing-related accomplishments reported. Responses were to an inventory asking about 12 different accomplishments, such as writing (and publishing) articles, technical manuals, speeches.
Readers' Evaluations of Student Writing Samples	
Writing sample evaluated according to (a) grade assigned by professor (b) holistic rating provided by specially trained readers	Study participants were asked to provide a typical example of writing they had done for a course. The request was for a paper of 10 or fewer pages drawn from the student's recent writing. Acceptable examples were essays, term papers, book reviews, lab reports, proposals, or case studies. (Writing samples were scored by readers using a 6-point holistic score scale.)
Personal statements	A personal statement, statement of goals, etc., that the student wrote as part of an application to graduate school. (Personal statements were scored by essay readers using a 6-point holistic score scale.)

Table 2

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

Background Characteristic	Study Sample (N = 470)	1991-92 GRE Test-Taking Population (N > 277,000)
GRE General Test		
GRE verbal mean (SD)	514 (120)	486 (118)
GRE quantitative mean (SD)	548 (128)	555 (139)
GRE analytical mean (SD)	563 (123)	537 (129)
Female (%)	65	54
Ethnicity %		
American Indian	1	1
Asian American	7	3
Black	11	7
Hispanic*	4	4
White	75	83
Other	2	2
English Best Language (%)	89	N.A.
Degree Objective (% Doctorate)	41	41
Undergraduate Major Field (%)		
Business	7	4
Education	7	9
Arts/Humanities	23	16
Social Sciences	25	22
Engineering	7	14
Natural Sciences	21	27
Other	11	9
Undergraduate Grade Average (%)		
A	12	12
A-	29	26
B	34	37
B-	18	17
C	7	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. Responses for the study sample are based on Ns of 394 to 470.

*Includes Mexican American and other Hispanic. There were no Puerto Rican examinees in the study sample.

Table 2

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

Background Characteristic	Study Sample (N = 470)	1991-92 GRE Test-Taking Population (N > 277,000)
GRE General Test		
GRE verbal mean (SD)	514 (120)	486 (118)
GRE quantitative mean (SD)	548 (128)	555 (139)
GRE analytical mean (SD)	563 (123)	537 (129)
Female (%)	65	54
Ethnicity %		
American Indian	1	1
Asian American	7	3
Black	11	7
Hispanic*	4	4
White	75	83
Other	2	2
English Best Language (%)	89	N.A.
Degree Objective (% Doctorate)	41	41
Undergraduate Major Field (%)		
Business	7	4
Education	7	9
Arts/Humanities	23	16
Social Sciences	25	22
Engineering	7	14
Natural Sciences	21	27
Other	11	9
Undergraduate Grade Average (%)		
A	12	12
A-	29	26
B	34	37
B-	18	17
C	7	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. Responses for the study sample are based on Ns of 394 to 470.

*Includes Mexican American and other Hispanic. There were no Puerto Rican examinees in the study sample.

Descriptive Statistics for the GRE Writing Exercises

The four "Issue" topics were comparable in terms of difficulty, with mean scores (sum of two readers) ranging from 7.7 to 7.9 (SDs from 2.0 to 2.2). As noted earlier, the majority of participants wrote two essays -- on either topics A and B or topics C and D. The correlation between first and second essays was .55. (The reader is reminded here that, by design, for each writer one topic was disclosed before the test and the other was not introduced until the test administration. Also, for each examinee one essay was written in 40 minutes and the other in 60 minutes. In all likelihood, these differences resulted in a lower correlation between essays than if both had been administered under exactly the same conditions.) The interreader correlations ranged from .77 to .80 for the four essay topics.

The means (sum of two readers) and standard deviations for the shorter exercises were as follows. For "Meaning of a Quotation" the mean was 4.3 (SD=1.5) on a 1 to 3 scale (sum of two readers); for "Specified Format", 8.0 (SD=2.5) on a 1 to 6 scale (sum of two readers); for "Revise a Composition", 3.2 (SD=1.1) on a 1 to 6 scale (only one reader was used for this exercise). The interreader correlations were .83, .94, and .83, respectively for these three exercises. (The estimate for the third exercise is based on a double reading of only a sample of responses.) These shorter measures related differentially to essay scores (r=.32, -.02, and .64, respectively), and none related very strongly to the others.

Descriptive Statistics for Writing Indicators

Comparison with Peers. Generally, study participants held themselves in high regard with respect to their writing skills. A majority reported themselves to be above average -- either somewhat above (48%) or well above (20%) -- when compared with peers in their major field of study. Another 25% regarded themselves as being about average, but only 6% felt they were either somewhat or well below average. This same evaluation was reported again about four months later by about one quarter (N=111) of the initial respondents. The correlation between early and later ratings was .70, suggesting moderately high reliability.

Evaluation of Performance on Writing Exercises. Study participants exhibited somewhat greater variation when asked to estimate how well they had performed on the writing exercises used in the study. Students rated their performances as follows:

Thorough and insightful ("6")	4%
Well developed ("5")	27%
Clearly competent ("4")	41%
Limited ("3")	23%
Weak or seriously deficient ("2" or "1")	6%

These frequencies appear to approximate a relatively normal distribution.

Success with Processes and Kinds of Writing. Table 3 shows the specific activities and kinds of writing about which participants were asked. As can be seen, respondents generally reported that they had experienced considerable success in writing. For each writing process and type of writing, a majority of subjects indicated that they had been either "quite successful" or "extremely successful." For each activity and kind of writing (with the exception of creative

writing), fewer than 10% of the participants reported that they had been "not at all" or "not very" successful. The same questions were asked about four months later of a subsample (N=111) of participants. The correlations between initial and follow-up responses for two scales -- "processes" and "kinds of writing" -- were .73 and .70, respectively, again suggesting a reasonably high level of reliability. (Internal consistency estimates based on coefficient alpha were .66 and .75.)

Grades and Accomplishments. Participants generally reported very high grades, both in courses that required writing (68% reported either an A or an A- average) and on their most recent assignment that required writing (80% reported either an A or an A- grade). Thus, there was relatively little variation among students on these particular indicators.

The degree to which students reported various writing-related achievements since high school is shown in Table 4. As expected, most of the 12 accomplishments listed on the survey were relatively rare. Only 5% of the respondents said that they had achieved 6 or more, and 33% reported none of the accomplishments. The mean number of accomplishments was 1.7 (SD=1.9).

Writing Samples and Personal Statements. More than 90% of those taking part in the study brought at least one writing sample with them to the testing center. Various features of the samples (when and where they were written, the amount of time spent writing them, and the grades that were awarded) are shown in Table 5. A slight majority of study subjects (54%) reported that the sample had been written more than a month prior to the testing date, but within the previous year. Most samples were written totally outside class. Some samples (20%) were produced in relatively short order (two hours or less), and others (34%) required significantly more time (ten or more hours). Despite our instructions for subjects to provide a typical example of their writing, most samples had received very favorable evaluations by professors, according to participants' reports. A majority (62%) reportedly received grades of A or A-, and significant numbers (29%) were not graded.

The essay readers trained for this study awarded, on average, relatively high scores to evaluate the papers that students brought with them (M=4.7, SD=1.1 on the 1-to-6 holistic scale that was used). A random sample of 22 papers was scored independently by two readers, yielding an interreader correlation of .68. This modest level of agreement suggests that, given the relatively minimal level of training provided for evaluating these samples, readers may have based their judgments on somewhat differing points of view. Given the highly diverse nature of the samples, it is perhaps surprising that a single, generic scoring rubric resulted in agreement that reached even this level.

Fewer than half of the study subjects provided personal statements, possibly because students regarded these documents as being of a more confidential nature than course-related samples. It seems more likely, however, that many students had not, at the time of testing, needed to write such statements. Of those providing personal statements, a significant proportion (37%) said that they had received at least some help in drafting their statements, and an even larger percentage (61%) acknowledged having had help in editing and revising them. Like the writing samples, personal statements were also judged, on average, to be of relatively high quality on the 6-point scale that was used (M=4.6, SD=0.9). Double scoring of a random sample of 29 statements yielded a reasonably high interreader correlation of .78.

Table 3

Reported Success in College with Various Writing Activities and Kinds of Writing

Activity or Kind of Writing	Mean[1]	SD	Percentage Reporting: Somewhat Successful (or less)	Quite Successful	Extremely Successful
Activity (Process)					
Thinking about an assignment (e.g., developing ideas, gathering information) .	4.2	0.7	15	50	35
Organizing (e.g., making outlines, deciding on the order to present ideas) . .	3.9	0.8	28	46	26
Revising (e.g., improving sentence phrasing, rearranging ideas, correcting grammar and punctuation)	3.8	0.9	33	44	23
Developing an effective writing style (e.g., expressing ideas clearly and in interesting ways)	3.9	0.9	33	43	24
Kind of Writing					
Personal writing (e.g., describing experiences or feelings as in a journal) . .	4.0	0.9	25	40	34
Creative writing (e.g., a poem or short story)	3.6	1.1	47	30	23
Persuasion (e.g., arguing a position or writing a letter to the editor)	3.8	0.8	38	42	20
Analysis/criticism (e.g., reviewing a book, movie, article, theory, work of art)	4.0	0.8	27	43	30
Description/explanation (e.g., describing an experiment on how something works) .	3.9	0.8	30	45	25
Examination writing (e.g., long essay answers)	3.8	0.9	34	44	22
Applied writing (e.g., preparing an application or resume)	3.9	0.8	27	48	25

Note. N = 466-469. For kinds of writing, results are based on only those who reported doing each kind of writing. Examinees who reported they had not done a particular kind of writing (N = 4 to 69 for the various activities/kinds) are not included in the calculations.

[1]Scale is 1 = not at all successful to 5 = extremely successful.

Table 4

Writing-Related Accomplishments Reported by Study Sample

Activity	Percentage Reporting Engagement
1. Wrote a feature article, column, or editorial that was published	28
2. Wrote technical manuals or other instructional material, for pay, for a company or other organization	23
3. Wrote advertising or public relations material, for pay, for a company or other organization	22
4. Wrote poetry, fiction, or essays that were published	21
5. Wrote a "letter to the editor" that was published	20
6. Was on the editorial staff of a publication or a radio or television station	15
7. Authored or coauthored a paper that was presented at a scientific meeting	13
8. Authored or coauthored an article that was published in a scientific journal	11
9. Wrote a speech for someone else that was given at a large public gathering (i.e., over 100 people)	10
10. Invited to participate in a writer's workshop sponsored by a statewide, regional, or national organization	6
11. Wrote a play that was publicly performed or a screenplay for a film that was publicly shown	3
12. Wrote the script for a dramatic or comedy show for radio or television that was publicly broadcast	2

Note. Percentages are based on 142 respondents.

Table 5

Characteristics of Writing Samples

Characteristic	Category	Percentage
When written	Within the month prior to testing	17
	One month to one year before testing	54
	More than a year prior to testing	29
Where written	Totally or mostly in class	5
	Mostly outside class	13
	Totally outside class	83
Time spent writing sample	Two hours or less	20
	Three to nine hours	46
	Ten or more hours	34
Grade on sample	Ungraded	29
	B or less	10
	A-	15
	A	47

Relationships among Nontest Writing Indicators

Table 6 displays the correlations among the various manifestations of writing skill collected for this study. As is apparent, the relationships among the indicators are clearly less than perfect, probably because these indicators either tap distinct aspects of a more general construct of writing proficiency or, if measuring a single aspect, reflect it with less than perfect reliability. Assuming the former, the best markers of a general construct appear to be comparisons with peers, grade on the most recent writing assignment, and self-reports of success with the processes of writing and with different kinds of writing. The median correlations of these indicators with all of the others in the set were .41 to .42. The most distinct indicators appear to be professor-assigned writing sample grades and quality of personal statements (median correlations with other indicators of .06 and .09, respectively.) Writing-related accomplishments were relatively unrelated to the other indicators also, exhibiting a median correlation of .13 with other indicators. The degree to which these weak relationships with other indicators reveal a specific contribution of these indicators to measuring writing proficiency (vs. simply a lack of reliability) is unclear.

Relationships of GRE Writing Exercises to Writing Indicators

Individual Indicators. Table 7 displays the correlations of each indicator of writing skill with each of the writing tasks investigated in the study. For the expository essays, two sets of correlations are reported -- one based on performance on both essays written by participants and the other based on the median correlation over first and second essays.

Correlations with essays were modest (mainly in the .30s) for each of the nontest writing indicators. The indicator having the highest correlation with essays was "comparison with peers"; the indicators exhibiting the lowest correlations were writing samples and personal statements. Correlations of indicators with two other exercises ("Meaning of a Quotation" and "Specified Format") were, with few exceptions, negligible. Two anomalies were significant ($p < .01$) correlations with reader-scored writing samples, and one significant ($p < .05$) correlation between "Meaning of a Quotation" and the personal statement. Performance on the "Revise a Composition" exercise bore modest relationships with the various indicators.[1]

In light of the extremely heterogeneous nature of the writing samples that were submitted, it seemed prudent to disaggregate them into more homogeneous, and therefore more comparable, subsets. Two classifications were used to define subsets: the amount of time students had spent writing the papers, and students' undergraduate major fields. The rationale for the first was that, compared with documents prepared over a long interval, samples written in a relatively short time period might relate more highly to timed essays written in a testing situation. Analyses did not, however, confirm this expectation: essay scores had the weakest relationships with samples reportedly prepared in an intermediate amount of time (3 to 9 hours). The correlations of essay scores with writing samples, by amount of time and type of evaluation were as follows:

Table 6

Correlations among Several Indicators of Writing Proficiency

Indicator	N	M	SD	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8a)	(8b)
1. Comparison with peers	465	3.8	0.8									
2. Success with processes	466	4.0	0.6	.55								
3. Success with kinds of writing	468	3.9	0.6	.53	.69							
4. Self-evaluation of study responses . . .	464	4.0	1.0	.45	.43	.46						
5. Grade average in writing courses	463	5.9	1.0	.50	.49	.45	.29					
6. Grade on most recent writing assignment	451	6.3	0.9	.41	.42	.42	.26	.54				
7. Accomplishments in writing	142	1.7	1.9	.25	.11	.24	.20	.16	.09			
8a. Writing sample (professor graded) . .	297	6.5	0.9	.24	.32	.26	.13	.34	.51	-.05		
8b. Writing sample (reader graded)	363	4.7	1.1	.06	.04	.08	-.04	.10	-.03	.13	-.04	
9. Personal statement	155	4.6	0.9	.20	.09	.09	.21	.13	-.04	-.08	-.01	.24

Note. Scales for variables were as follows:

Variable 1 : 1=well below average to 5=well above average.
Variable 2 : Average of four activities on a scale of 1=not at all successful to 5=extremely successful.
Variable 3 : Average of seven kinds of writing on a scale of 1=not at all successful to 5=extremely successful
Variables 4, 8b, 9 : 1=seriously deficient in writing skills to 6=well-developed response demonstrating mastery of the elements of effective writing.
Variables 5, 6, 8a : 1=D or less to 7=A.
Variable 7 : Sum of yes responses to 12 different writing accomplishments.

Correlations are based on samples of from N=48 (personal statements with accomplishments), N=83 (writing sample -- professor graded with personal statements), N=103 (writing sample -- reader graded with personal statements), ... to N=466 (for several pairs of variables).

For Ns of 48, 83, and 103, correlations of .29, .22, and .19 are significant at the .05 level, and correlations of .37, .28, and .25 are significant at the .01 level, two-tailed test.

Table 7

Correlations of Alternative Writing Exercises
with Indicators of Writing Proficiency

Indicator (Ns)	Writing Exercise: Essay	Meaning of a Quotation (MQ)	Specified Format (SF)	Revise a Composition (RC)
1. Comparison with peers (N = 299, 106, 115, 170)	.46/.41	.19	.16	.29
2. Success with processes (N = 300, 106, 115, 169)	.32/.28	.15	.14	.24
3. Success with kinds of writing (N = 301, 107, 115, 170)	.36/.32	.21	.08	.38
4. Self-evaluation of study responses (N = 300, 106, 113, 168)	.39/.34	.09	.05	.24
5. Grade average in writing courses (N = 298, 106, 113, 168)	.35/.31	.23	.18	.31
6. Grade on most recent writing assignment (N = 290, 104, 111, 163)	.38/.34	.07	.19	.25
7. Accomplishments in writing (N = 141, 32, 23, 32)	.34/.30	.04	-.04	.03
8a. Writing sample (professor graded) (N = 185, 71, 76, 114)	.18/.16	-.08	.05	.23
8b. Writing sample (reader graded) (N = 236, 86, 85, 131)	.18/.15	.29	.49	.14
9. Personal statement (N = 105, 31, 38, 58)	.17/.15	.37	.11	.21
MEDIAN	.35/.31	.17	.13	.24

Note. First number in the pair under essay is the correlation of the indicator with the sum of two essays. The second is the median of correlations with first and second essays written.

For Ns of 30, 100, and 300, correlations of .36, .20, and .11 are significant at the .05 level, and correlations of .46, .25, and .15 are significant at the .01 level, two-tailed test.

Hours	Reader's Score	Professor's Grade
Two or fewer	.29*	.22
3 to 9	.13	.12
10 or more	.36**	.28*

*p<.05, two-tailed, **p<.01, two-tailed

Relationships also varied somewhat by major field of study. The correlations, by major field of study, between essay scores and writing sample scores (for essay readers and professors) were as follows:

Major Field	Reader's Score	Professor's Grade
Arts/humanities	.45**	-.06
Social sciences	.43**	.25
Natural sciences/engineering	.06	.55**
Business/education/other	.09	.13

(Sample sizes were insufficient for computing comparable correlations by major field for the shorter writing exercises.)

Composite Indices. Several indices of writing proficiency were constituted on the basis of logically related subsets of indicators, as classified earlier in Table 1. Each individual indicator was first z-scaled to a mean of 0, standard deviation of 1. The resulting z scores were then added to form several composite indices, which were differentially related. Index 3, based on evaluations of writing samples, exhibited negligible correlations with the other indices (r = -.07 to .13). Index 1, a composite of self-estimates, correlated moderately with each of the three self-report indices (r = .55 to .64). (The three self-report indices were highly related to one another, as they shared common components.)

The correlations of the alternative writing exercises with each of these indices are given in Table 8. This table shows moderate correlations of essay scores with the various indices -- correlations that are somewhat higher than those based on individual indicators. Though available for a relatively small subsample of study participants, an index based on self-reported grades and writing accomplishments bore the strongest relationship to essay-based scores. Again, the essay task seems to have the most potential, and the "Revise a Composition" exercise the next most promise.

Relationships of GRE General Test Scores to Writing Indicators and Essays

Table 9 features the correlations of each nontest writing indicator with scores from each portion of the GRE General Test. Overall, GRE verbal scores bear a stronger relationship to these indicators than do either quantitative or analytical scores. Comparing these findings with those in Table 7, it appears that essay scores are, on average, slightly more highly related to the various indicators of writing than are any of the scores from the current GRE General Test.

Table 8

Correlations of Writing Exercises
with Alternative Indices of Writing Proficiency

Index (Ns)	Writing Exercise: Essay	Meaning of a Quotation (MQ)	Specified Format (SF)	Revise a Composition (RC)
1. Self estimates Processes, kinds of writing, and comparison with peers (N = 299, 106, 115, 169)	.44/.39	.22*	.15	.41***
2. Self reports				
a. Grades in writing courses, most recent writing grade (N = 290, 104, 111, 163)	.42/.37	.17	.19*	.31***
b. Grades in writing courses, most recent writing grade, accomplishments in writing (N = 136, 30, 23, 31)	.59/.53	.19	.11	.40
c. Grades in writing courses, most recent writing grade, professor grade on writing sample (N = 184, 71, 75, 114)	.36/.31	.14	.13	.39***
3. Readers' evaluations Reader score on writing sample and on personal statements (N = 71, 20, 24, 40)	.31/.27	.22	.41*	.26

Note. All correlations involving the essay exercise are significant at either the .01 or the .001 level. The significance of other correlations is indicated as follows.
*$p<.05$, **$p<.01$, ***$p<.001$.

Table 9

Correlations of Writing Indicators with GRE General Test Scores

Indicator	GRE Scores		
	Verbal	Quantitative	Analytical
1. Comparison with peers	.43	.04	.20
2. Success with processes	.22	-.06	.10
3. Success with kinds of writing	.26	-.08	.14
4. Self-evaluation of study responses	.24	.01	.16
5. Grade average in writing courses	.36	.06	.28
6. Grade on most recent writing assignment	.30	-.00	.23
7. Accomplishments in writing	.17	-.26	-.03
8a. Writing sample (professor graded)	.18	-.03	.05
8b. Writing sample (reader graded)	.16	.10	.18
9. Personal statement	.23	.24	.25
MEDIAN	.24	.01	.17

Note. Ns=449 to 466 for indicators (1) - (6).
Ns=142, 295, 361, and 155 for indicators (7) - (9), respectively.

Students' scores on the essays written for this study also correlated with GRE scores. The correlations between a single essay (median over both essays written) and GRE verbal, quantitative, and analytical scores were .53, -.01, and .30, respectively. The corresponding correlations for the "Revise a Composition" exercise were .63, .09, and .49. (The "Meaning of a Quotation" exercise was much less strongly related to GRE scores, r = .29, .14, and .19, and the "Specified Format" exercise was entirely unrelated to GRE scores.) There is, therefore (as expected), some overlap with scores from the current GRE General Test for both the essay and the revision exercises. An issue here, then, is the extent to which these exercises tap more general (verbal) reasoning skills in addition to writing proficiency. These figures are not so high, however, as to preclude a unique contribution by a writing measure. By comparison, GRE Subject Test scores, which have been shown to contribute uniquely to the prediction of academic performance when used in combination with GRE General Test scores, correlate at least as highly with the General Test (Briel, O'Neill, & Scheuneman, 1993). For example, in 1990-91 the correlations for the Subject Test (Physics) that was least highly related to the General Test were .18, .59, and .36 with GRE verbal, GRE quantitative, and GRE analytical scores, respectively. Comparable correlations were .79, .70, and .69 for the Subject Test in Sociology, the test most strongly related to General Test scores. More typical of correlations were those reported for the Economics Subject Test (.52, .66, and .56).

Discussion

Several writing tasks were administered to volunteers recruited from the GRE test-taking population. Volunteers represented the population quite well in terms of ability level and demographic characteristics. Study participants reported a variety of other information about their writing skills and experiences with writing, and most submitted a course-related sample of writing and/or a personal statement written for application to graduate school.

Major Findings

Among the major findings were the following:

- The various nontest indicators of writing skills were weakly to moderately interrelated, suggesting either that they embody different aspects of writing proficiency or that they are only moderately reliable. Very probably, each indicator also reflected somewhat different extraneous influences.

- The writing exercises that were investigated related differentially to the various nontest indicators. Each of two shorter exercises requiring examinees to provide their reflections on a quotation and to write in a specified format exhibited only very weak relations with the indicators. A third 20-minute exercise, involving the revision of a poorly written composition, performed significantly better (and nearly as well as the longer essay). The expository essay task, however, appeared to be the most promising of the lot, generally, with respect to its correlations with other manifestations of subjects' writing skills. Still, however, correlations with student-supplied writing samples were weak. This latter finding corresponds to results of the National Assessment of Educational Progress comparisons of classroom writing (of fourth- and eighth-grade students) with writing performed under strictly timed conditions (Gentile, Martin-Rehrmann, & Kennedy, 1995).

- Although the quality of writers' essays was not especially strongly related to any of the nontest indicators, it did relate modestly to most of them. Moreover, essay-based scores were somewhat more highly related to several composite indices of writing proficiency based on sets of logically related individual indicators than they were to individual indicators.

- Performance on expository essays was, on balance, related somewhat more strongly to alternative indicators of writing than were GRE General Test scores, suggesting that a writing test comprised of essay topics would contribute to the current battery of GRE measures.

- An exercise involving the revision of a poorly written essay might also be worthy of further exploration. It is noteworthy that this 20-minute exercise, scored in our study by only a single reader, bore significant correlations with several indicators of writing skill. Had this exercise been given "equal time" with the 40- and 60-minute essays, it might have fared even better. Furthermore, performance might have been more impressive still if, instead of using holistic scoring methods designed primarily for expository essays, we had employed the more detailed (but time-consuming) procedures developed by Godshalk, Swineford, and Coffman (1966) and studied by Breland, Bonner, and Kaplan (1993).

Study Limitations

The limitations of the study apply principally to the nature of the nontest writing criteria that were used to gauge the meaning of the alternative writing exercises. First, no claim is made that, as a collection, the assortment of evidence assembled here constitutes an especially compelling theoretical framework that adequately spans the construct of writing proficiency. Each of the indicators has its own weaknesses, some of which are shared with others; however, each contributes, we believe, in at least some small way to the appraisal of writing skills.

Comparisons with peers and other self-assessments, for example, appear to have significant validity (as gauged by correlations with other objective indicators of ability) in a number of different contexts (Mabe & West, 1982), particularly in academic settings (e.g., Baird, 1976), and most especially in self-assessments of foreign language skills (Blanche & Merino, 1989). As Upshur (1975) noted with regard to self-assessments of language skills, learners often have access to the complete spectrum of their successes and failures, whereas other indicators may represent a much smaller segment of experiences. Self-evaluations suffer, however, from their somewhat dubious objectivity: although individuals may sometimes be the best judges of their own capabilities, they are undoubtedly in many circumstances also the least able. Evidence that most people regard themselves as "above average" is not confined to Garrison Keillor's fabled Lake Wobegon: there was ample proof of this phenomenon in our data. Moreover, self-assessments may have differing bases from person to person, and they may be influenced more heavily by some characteristics (e.g., those that are more readily observed) or events (e.g., the more vivid or unusual ones) than by others. In addition, personal traits such as modesty, self-confidence, and defensiveness may color the appraisal and reporting of one's own competencies.

Self-reports involving more factual matters may sometimes be suspect also. Contamination may result from forgetting (or selective remembering), responding according to

socially accepted norms, especially for some kinds of information, or deliberate falsification. Nonetheless, such reports have proved to be of sufficient accuracy for a wide variety of purposes. This applies in particular to reports of grades and other academic accomplishments (Baird, 1976; Freeberg, 1988; Freeberg, Rock, and Pollack, 1989; Laing, Sawyer, and Noble, 1987; Sawyer, Laing, and Houston, 1988).

The writing samples that we collected have their own characteristic strengths and limitations as validation criteria, both in terms of the samples themselves and the manner in which they were evaluated. A concern with the course-related samples, for instance, is their lack of standardization: the samples we studied were, obviously, written for a variety of different purposes and audiences. A similar concern applies, to a somewhat lesser degree, to personal statements. A potentially greater problem with personal statements, however, at least as samples of writing, is the uncertainty that often accompanies their authorship. A majority of the participants in our study acknowledged that they had obtained help at least in editing and revising their statements, and many said they had received assistance with their initial drafts as well.

Writing samples based on professors' assignments, on the other hand, although potentially the most compelling source of validity evidence, suffered still other shortcomings. As mentioned earlier, although study participants were asked to submit typical samples of their writing, they seemed to furnish the best examples (as suggested by the high grades that most samples had received). As expected, the samples formed a heterogeneous collection, ranging from an engineer's memorandum on "results of a study of centrifugal pumps" to a paper on "applying neo-Aristotelian criticism to the beverage container recycling effort in California" to a personal account of "the life and times of an exotic dancer." To the best of our knowledge, this effort represents the first time that holistic scoring has been applied to such a variety of disparate papers from various academic fields of study. Not surprisingly, then, some readers did not feel fully confident, at least initially, in judging all of these pieces. Moreover, the holistic scoring rubric used in the study did not fit all samples equally well. Stretching an "all-purpose" rubric to encompass such a wide variety of tasks surely reduced the reliability of evaluations.

We were able to gather some rudimentary information on the nature of the writing samples -- for example, when they were written and how much time was involved in writing them. In retrospect, however, other, uncollected information might have been even more useful in classifying the wide variety of writing samples submitted, and in identifying specific circumstances in which the relationship between the judged quality of test essays and that of writing samples would have been stronger. Information about the objectives of the assignment, the directions that were given, the specific requirements (e.g., page length), and whether the assignment was chosen by the student or specified by the teacher might have been constructive. Some quite elaborate schemes have been devised for classifying writing assignments (e.g., Purves, Soter, Takala, & Vahapassi, 1984), and the application of these might have proved informative. Breaking out the study sample by undergraduate field did suggest that these relationships were greater in some fields (arts/humanities and the social sciences) than in others.

In addition, students reported the grades that their professors had assigned to the samples. However, the shortcomings of academic grades -- for example, that they reflect differential standards both among courses and among professors -- are well known. Further, the

grades assigned to the samples provided by study participants were undoubtedly based on characteristics other than (or in addition to) the quality of students' writing, and they were quite restricted in their range, thus precluding very high correlations with any other variables.

Conclusion

Despite the myriad qualifications detailed above, the study does, we believe, constitute a reasonable beginning to accumulating the kind of evidence required to establish the validity of a GRE writing test for facilitating the selection of prospective graduate students. Indeed, even with the limitations of the criterion indicators, two of the exercises investigated here had quite respectable relations with composite indices of writing skill. The study results also provide clues as to alternative methods of investigation and additional kinds of evidence that might be pursued in any future efforts.

One kind of evidence that should be gathered eventually, we suggest, is information of a more prospective nature. All of the information gathered here was either concurrent with or retrospective to the writing that subjects completed for this study. Confirmation needs to be provided in terms of the relationship of GRE writing exercises to students' success in meeting the writing demands of graduate study. In addition, a more focused sampling of schools and departments within schools, instead of the broad-based representative sampling that was elected for this study, would help in collecting more homogeneous samples of course-related writing as validation criteria.

Endnotes

1. As noted earlier, in contrast to other exercises, the "Revise a Composition" exercise was read by only one reader, thus resulting in less reliable scores for this measure and lower correlations with other variables than those for the exercises that received a double reading. The interreader correlation for a sample of the RC exercises that were read twice was .94. In the interest of a fairer comparison with the other exercises, we used the Spearman-Brown formula to estimate the reliability of two readers for this measure (.97). Further, had two readers been employed instead of one, we estimate that the correlations of variables with RC would have increased by at most .01.

Again in the interest of fair comparison among the alternative exercises, we computed the actual correlations of a single essay, read only once, with other variables. Correlations with the nontest indicators decreased slightly -- from .00 to .03 for the various indicators.

References

Baird, L. L. (1976). Using self-reports to predict student performance. New York: College Entrance Examination Board.

Baird, L. L. (1979). Development of an inventory of documented accomplishments for graduate admissions (GRE Board Report No. 81-1R). Princeton, NJ: Educational Testing Service.

Blanche, P., & Merino, B. (1989). Self-assessment of foreign language skills: Implications for teachers and researchers. Language Learning, 39, 313-40.

Breland, H. M. (1983). The direct assessment of writing skill: A measurement review (College Board Report No. 83-6, ETS RR No. 83-32). New York: College Entrance Examination Board.

Breland, H. M., Bonner, M. W., & Kaplan, R. M. (1993). Investigation of a computerized constructed response item type for writing assessment (Proposal to the Graduate Record Examinations Board). Princeton. NJ: Educational Testing Service.

Breland, H. M., Camp, R., Jones, R. J., Morris, M. M., & Rock, D. A. (1987). Assessing writing skill. New York: College Entrance Examination Board.

Breland, H. M., & Gaynor, J. (1979). A comparison of direct and indirect assessments of writing skill. Journal of Educational Measurement, 16, 119-128.

Briel, J. B., O'Neill, K. A., & Scheuneman, J. D. (Eds.) (1993). GRE technical manual: Test development, score interpretation, and research for the Graduate Record Examinations program. Princeton, NJ: Educational Testing Service.

Cooper, P. L. (1984). The assessment of writing ability: A review of research (GRE Board Research Report No. 82-15R, ETS RR No. 84-12). Princeton, NJ: Educational Testing Service.

Enright, M. K., & Gitomer, D. (1989). Toward a description of successful graduate students. (GRE Board Report No. 85-17R). Princeton, NJ: Educational Testing Service.

Flower, L. S., & Hayes, R. J. (1980). The dynamics of composing. In L. W. Gregg and E.R. Steinberg (Eds.), Cognitive processes in writing (pp. 31-50). Hillsdale, NJ: Erlbaum.

Freeberg, N. E. (1988). Analysis of the revised student descriptive questionnaire, Phase I (College Board Report No. 88-5, ETS RR No. 88-11). New York: College Entrance Examination Board.

Freeberg, N. E., Rock, D. A., & Pollack, J. (1989). Analysis of the revised student descriptive questionnaire: Phase II (College Board Report No. 89-8, ETS RR No. 89-49). New York: College Entrance Examination Board.

Gentile, C. A., Martin-Rehrmann, J., & Kennedy, J. H. (1995). Windows into the classroom: NAEP's 1992 writing portfolio. Washington, DC: U.S. Department of Education.

Godshalk, F. I., Swineford, F., & Coffman, W. E. (1966). The measurement of writing ability (Research Monograph No. 6). New York: College Entrance Examination Board.

Hartnett, R. T., & Willingham, W. W. (1980). The criterion problem: What measure of success in graduate education? Applied Psychological Measurement, 4, 281-291.

Hout, B. (1994). An introduction to assessing writing. Assessing Writing, 1, 1-9.

Huddleston, E. (1954). Measurement of writing ability at the college-entrance level: Objective vs. subjective testing techniques. Journal of Experimental Psychology, 22, 165-213.

Laing, J., Sawyer, R., & Noble, J. (1987). Accuracy of self-reported activities and accomplishments of college-bound students (ACT Research Report Series 87-6). Iowa City, IA: The American College Testing Program.

Mabe, P. A. III, & West, S. G. (1982). Validity of self-evaluation of ability: A review and meta-analysis. Journal of Applied Psychology, 67, 280-296.

Miller, M. D., & Crocker, L. (1990). Validation methods for direct writing assessment. Applied Measurement in Education, 3, 285-296.

Powers, D. E., & Fowles, M. E. (in press). Effects of pre-examination disclosure of essay topics (GRE Board Report 93-26A).

Purves, A. C., Soter, A., Takala, S., & Vahapassi, A. (1984). Towards a domain-referenced system for classifying composition assignments. Research in the Teaching of English, 18, 385-415.

Reilly, R. R. (1976). Factors involved in graduate student performance. American Education Research Journal, 13, 125-138.

Sawyer, R., Laing, J., Houston, M. (1988). Accuracy of self-reported high school courses and grades of college-bound students (ACT Research Report Series 88-1). Iowa City, IA: The American College Testing Program.

Upshur, J. (1975). Objective evaluation of oral proficiency in the ESOL classroom. In L. Palmer and B. Spolsky (Eds.), Papers on language testing 1967-1974. Washington, DC: TESOL.

Wainer, H. (1993). Measurement problems. Journal of Educational Measurement, 30, 1-21.

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

Wiggins, G. P. (1993). Assessing student performance: Exploring the purpose and limits of testing. San Francisco: Jossey-Bass.

Appendix A

GRE Writing Measure Study Questionnaire

About the exercises you were given to write on today

1. Did you have enough time for: *(Circle one number for each.)*

	Yes, more than enough	Yes, just enough	No, not quite enough	No, not nearly enough	I did not get an exercise of this length
A. The 20-minute exercises	4	3	2	1	0
B. The 40-minute essay . . .	4	3	2	1	0
C. The 60-minute essay . . .	4	3	2	1	0

2. Overall, how would you rate the writing you did for the tasks you were given?
(1 = lowest to 6 = highest)

Seriously deficient 1
Weak 2
Limited 3
Clearly competent 4
Well developed 5
Thorough and insightful 6

About preparing for this test

3. How did you prepare for this test? *(Circle one number for each statement.)*

	No	Yes
A. I spent time thinking generally about the topics that were sent	1	2
B. I thought about specific points or concrete examples to cover for the topics	1	2
C. I "brainstormed" about ideas	1	2
D. I did some reading about the topics that were sent	1	2
E. I wrote an outline for one or both topics	1	2
F. I drafted a sample essay for one or both topics	1	2
G. I asked someone to read and comment on my essays	1	2
H. I read the information on how the essays are scored	1	2
I. I read the exemplary essays you sent	1	2
J. Other (please describe below)	1	2

Other: _____

About the writing samples (report, etc.) that you brought with you

4. When did you write (i.e., complete) it?

Within the past month 1
Two to four months ago 2
Five to twelve months ago 3
More than one year ago 4
I did not bring a sample ..... 0 (Skip to Question 8)

5. Was it written in or outside class?

Totally in class 1
Mostly in class 2
Mostly outside class 3
Totally outside class 4

6. About how much time from start to finish did you spend writing it?

Less than 1 hour 1
About 1 to 2 hours 2
About 3 to 9 hours 3
About 10 to 20 hours 4
More than 20 hours 5

7. What grade did you receive on it?

A	7	C	3
A-	6	C-	2
B	5	D or less	1
B-	4	Was not graded	0

8. Regarding the personal statement you provided, how much help did you get in drafting, revising, and editing it? *(Circle one number for each.)*

	None	Very little	A moderate amount	A substantial amount
A. Drafting	1	2	3	4
B. Editing/revising	1	2	3	4

I did not provide a personal statement 0

About your writing

9. 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 1
Somewhat more slowly 2
About the same 3
Somewhat more quickly 4
Much more quickly 5

10. 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 1
Usually 2
Sometimes 3
Seldom 4
Hardly ever 5

11. How do you think your writing compares with that of other students in your major field?

Well above average 5
Somewhat above average 4
About average 3
Somewhat below average 2
Well below average 1

About your experiences with the process of writing

12. For each of the activities listed, please indicate how successful you have been in your college courses. *(Circle one number for each.)*

5 = extremely successful 3 = somewhat successful 1 = not at all successful
4 = quite successful 2 = not very successful

	Extremely Successful	Quite Successful	Somewhat Successful	Not very Successful	Not at all Successful
A. Thinking about an assignment (e.g., developing ideas, gathering information)	5	4	3	2	1
B. Organizing (e.g., making outlines, deciding on the order to present ideas)	5	4	3	2	1
C. Revising (e.g., improving sentence phrasing, rearranging ideas, correcting grammar and punctuation .	5	4	3	2	1
D. Developing an effective writing style (e.g., expressing ideas clearly and in interesting ways)	5	4	3	2	1

About completing various writing assignments

13. For each of the kinds of writing listed, please indicate how successful you have been in your college courses. *(Circle one number for each.)*

	Extremely successful	Quite successful	Somewhat successful	Not very successful	Not at all successful	I did not do this
A. Personal writing (e.g., describing experiences or feelings as in a journal) . .	5	4	3	2	1	0
B. Creative writing (e.g., a poem or short story)	5	4	3	2	1	0
C. Persuasion (e.g., arguing a position or writing a letter to the editor)	5	4	3	2	1	0
D. Analysis/criticism (e.g., reviewing a book, movie, article, theory, work of art)	5	4	3	2	1	0
E. Description/explanation (e.g., describing an experiment or how something works)	5	4	3	2	1	0
F. Examination writing (e.g., long essay answers)	5	4	3	2	1	0
G. Applied writing (e.g., preparing an application or resumé)	5	4	3	2	1	0

About your grades

14. Approximately, what grades or grade averages did you get (A) overall in college, (B) in courses that required at least some writing, and (C) on the most recent assignment for which you had to write a report, essay, etc. (If your college does not use letter grades, please mark the grade that is the closest equivalent.) *(Circle one number for each.)*

Grade	(A) Overall average	(B) Courses requiring writing	(C) Most recent writing assignment
A	7	7	7
A-	6	6	6
B	5	5	5
B-	4	4	4
C	3	3	3
C-	2	2	2
D or less	1	1	1

15. May we use, as examples in our research reports, samples of your writing? You will not, of course, be identified by name. *(Circle one number for each.)*

	No	Yes
A. Personal statement	1	2
B. Writing sample you brought with you	1	2
C. Essays you wrote today	1	2

Your signature _____

Social Security Number (for payment purposes) ___-___-____

Comments

If you have any comments about this study or about the writing exercises, please make them here.

_____

_____

_____

_____

Thanks for your help.

54020-14302 • Y126P.5 • 209240