GRE®

RESEARCH

Further Validation of a Writing Assessment for Graduate Admissions

Donald E. Powers
Mary E. Fowles
Cynthia K. Welsh

July 1999

GRE Board Research Report No. 96-13R

ETS Research Report 99-18

Educational Testing Service, Princeton, New Jersey 08541

Further Validation of a Writing Assessment for Graduate Admissions

Donald E. Powers
Mary E. Fowles
Cynthia K. Welsh

GRE Board Report No. 96-13R

July 1999

This report presents the findings of a
research project funded 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.

Educational Testing Service
Princeton, New Jersey 08541

Acknowledgments

We'd like to thank Dr. Brian Huot, Professor of English at the University of Louisville; RoseAnn Morgan, Professor of English, Middlesex College, New Jersey; Dr. Art Young, Professor of English and Engineering at Clemson University, and Dr. Agnes Yamada, Professor of English and Department Chair at California State University at Dominguez Hills. They helped us to develop the scoring methods used to evaluate participants' writing samples and provided invaluable advice on several other aspects of the research.

Special thanks go also to the following ETS staff members: Joi Bracy-Strachn for retrieving information from GRE files; Norma Norris and Inge Novatkoski for analyzing our data; Jackie Briel and Rob Durso for their advice and encouragement; Gary Schaeffer for helping us obtain and match data from the Part I study; Cindy Evans for assistance with managing data files and payment to participants; Ruth Yoder for preparing tables, graphs, and the final report, as well as for conducting a wide variety of tasks throughout the course of the project; and Hunter Breland and Claudia Gentile for helpful reviews of the draft report.

We also acknowledge the support of the GRE Board and its Research Committee.

Finally, special thanks go to the faculty members who read and evaluated some 4,000 writing samples and to the students who provided them.

Abstract

The objective of the study reported here was to collect validity evidence for a proposed writing assessment for graduate admissions -- the Graduate Record Examinations (GRE®) Writing Assessment. In particular, the objective was to investigate the relationship between student performance on each of two exercises being considered for the assessment and several nontest indicators of writing skill and achievement, and to thereby establish the degree to which performance on the GRE Writing Assessment is related to writing performance in academic settings. A variety of nontest indicators were examined, but a particular focus was the quality of students' course-related writing samples. Two such writing samples were collected for each participant in the study, as was considerable information about the nature of these samples. These data enabled an estimate of the reliability and generalizability of writing samples as a validity criterion. The data also permitted an analysis of the conditions and circumstances under which performance on course-related writing assignments and performance on the two new writing exercises relate to one another, thus facilitating the interpretation of scores derived from the new writing assessment. The results revealed modest relationships between performance on the writing assessment essays and various nontest indicators of writing ability. Performance on the GRE Writing Assessment exhibited the strongest relationship with course-related writing samples, arguably the most compelling of the nontest indicators. There was no indication that the relationship between the essays and course-related writing samples may depend on particular characteristics of the sample.

KEY WORDS:
validity
writing assessment
graduate admissions
Graduate Record Examinations
test validation

Table of Contents

List of Tables

List of Figures

Introduction

Test validation -- determining the appropriateness, meaningfulness, and usefulness of test score inferences -- is a continuing and possibly never-ending process. According to Messick (1989), although relevant evidence can take many forms, in practice there are but a few distinct kinds of validity data. One major category of validity evidence has been dubbed variously as (a) relations to external variables (Messick, 1989), (b) correlations with practical criteria (Cronbach, 1990), and (c) associations with variables that test scores should and should not relate to (Hambleton & Rogers, 1991). It is this category of practical, external variables that was the focus of both the current study and a previous, related study (Powers, Fowles, & Boyles, 1996). The purpose of both studies was to collect validity evidence for a proposed writing assessment for graduate admissions -- The Graduate Record Examination (GRE®) Writing Assessment.

The GRE Writing Assessment is designed to provide evidence of examinees' ability to:

- articulate complex ideas clearly and effectively
- examine claims that accompany evidence
- support ideas with relevant reasons and examples
- sustain a well-focused, coherent discussion
- accurately apply the conventions of standard written English such as, grammar, usage, and mechanics

A specific objective of this study was to investigate the relationships between the writing skills that can be elicited in an assessment situation and the skills that college and university students need in order to demonstrate their learning. Also of interest was the extent to which these relationships might differ according to the demographic and background characteristics of writers, the characteristics of their academic writing, and the circumstances under which academic writing is performed. According to a comprehensive review provided by Huot (1990), an interest in validity issues (as we have defined them here) has not been one of the prevailing concerns in the assessment of writing. We feel that it should be.

To accomplish the study objectives, we examined the relationships between holistic judgments of the quality of students' course-related writing samples and their scores on the two analytical writing tasks that will comprise the GRE Writing Assessment -- "Analyze an Argument" and "Present Your Perspective on an Issue." Relationships between scores on these two GRE writing tasks and several other nontest indicators of student writing skill, such as self-assessments (e.g., comparisons with peers) and self-reports of various writing-related accomplishments and achievements (e.g., grades in undergraduate courses that required writing) were also examined.

How the Current Study Builds on Previous Research

For the prior (Powers et al., 1996) investigation, data were collected from a sample of approximately 400 undergraduate students, all of whom had taken the GRE General Test and who, as a group, were similar to the entire GRE test-taking population. At that time, subsamples of study participants

tried various, alternative writing exercises that were being considered for possible inclusion in a GRE writing assessment. All study participants were asked to provide information about several nontest indicators of their writing skill and to submit a single course-related writing sample.

Analyses revealed that the various writing exercises exhibited distinctly different patterns of correlations with the nontest indicators. On the basis of these correlational patterns, a task requiring students to take a stand on an issue was judged to be the most promising of the various exercises being considered for the test. (The "Analyze an Argument" writing exercise was not included in the 1996 study by Powers et al. because, at that time, the GRE Writing Advisory Committee had not yet approved this task for pilot testing.) None of the other experimental exercises proved to be as defensible as the Issue task, and so they were eliminated from further consideration.

For the Issue task, holistic scores correlated modestly (mainly in the .30s) with each of the nontest indicators of writing skill. The sole exception was a very weak correlation with the quality of students' course-related writing samples. (Both the Issue essays and the course-related writing samples were evaluated on a 6-point holistic scale by specially-trained university faculty). Because performance on course-related writing samples was probably the most compelling of any of the indicators examined, this very weak correlation was disappointing.[1]

Powers et al. (1996) hypothesized that the unexpectedly low correlations between GRE essays and students' course-related writing samples were largely a function of the way in which the writing samples were solicited, and to some degree, evaluated. Because study participants were given little guidance as to the kinds of writing samples or course-specific information to submit, the resulting writing samples constituted an extremely heterogeneous lot. In an effort to achieve more homogeneous subsets, the authors classified the writing samples on the basis of students' undergraduate major fields and the amount of time they had spent writing their papers. The rationale for the first classification was that, because writing requirements vary by academic discipline, so too might their relationship to GRE essays. The rationale for the second classification was that, compared with documents written over a long interval, samples written in a relatively short time period might relate more highly to timed essays composed in a testing situation. The relationships between essay performance and nontest indicators did in fact depend on both of these

---

[1] The other results of our study compared favorably with earlier, related research, as documented by Breland (1983) and by Miller and Crocker (1990), for instance. 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 to12 months later (reported for two studies) were in the .50s. Later grades, usually based on only courses in English, correlated in the .20s and low .30s.

A study by Breland, Camp, Jones, Morris, and Rock (1987) 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. This study was characterized by Miller and Crocker (1990) as "the most comprehensive criterion-related validity study [of a direct measure of writing]" (p. 293).

Most recently, Breland, Kubota, and Bonner (in press) collected as a criterion measure four take-home writing assignments (both drafts and final papers) from first-year undergraduates. The SAT II essay was moderately predictive of performance on this criterion (r = .39), of English course grades (r = .35), and of a self-report composite (r = .38).

Breland, Bridgeman, and Fowles (in press) have provided further review of the validity of writing assessments when various kinds of criteria -- grades, teachers' ratings, and writing performance -- have been used.

variables. Essay scores related most strongly to course-related assignments in the arts and humanities and in the social sciences, and (counter-intuitively) to writing samples on which students had spent 10 or more hours.

Thus, rudimentary information on the nature of the writing samples -- that is, the field of study in which they were written and how much time was involved in writing them -- proved informative. The authors reasoned, in retrospect, that collecting additional information about other characteristics of students' writing samples might be useful for identifying the specific circumstances or conditions that determine the strength of the relationship between the quality of GRE essays and these samples. A major feature of the study reported here, therefore, was the collection of considerably more information about the nature of students' writing samples. Of interest, for example, were the objectives of the assignment, the specific requirements (such as the number of pages), and who chose the topic -- the student or the teacher.

In addition to collecting more information about the students' writing samples, we extended the previous study by:

- involving a larger sample of participants, thus enabling a comparison of predictor-criterion relationships for several subgroups of examinees -- defined by gender, ethnicity, English language background, and undergraduate field of study, for example -- in order to ascertain the validity of the assessment for subgroups

- investigating performance on both the Argument and Issue assessment exercises in order to assess the validity of each task for different academic fields and for different kinds of writing demands

- obtaining two course-related writing samples from each student, so that the reliability/generalizability of writing samples as a criterion could be assessed

- including a measure of examinee motivation so that we could assess the effects of participants' effort on estimates of the validity of the writing assessment

- facilitating additional research on the GRE Writing Assessment by ascertaining students' interest in future participation and by obtaining addresses where they could be reached

The overarching objective of the present study was to establish the degree to which the performance of college and university students on the GRE Writing Assessment reflects their writing performance in an academic setting.

Method

Sample Selection

The present study is the second part of a two-part research project. For this reason, we were dependent on the sample selected in part one of the study by Schaeffer, Briel, and Fowles (in preparation). Schaeffer et al. recruited study participants (college juniors and seniors and first-year graduate students) in the fall of 1997 from a total of 26 geographically diverse colleges and universities, including several with relatively large percentages of minority students. Through a variety of means (flyers, posters, and campus news releases, for instance), potential participants were informed about a two-part research study for a

proposed GRE Writing Assessment. In part one, students were required to write either two Issue essays, two Argument essays, or one of each, and to complete a brief background questionnaire that included questions about their reactions to the writing assessment. Further details about sample selection and about part one are provided in the report by Schaeffer et al. Students were eligible for the second part of the study (the validity study reported here) only if they participated in the first part. Although they were not required to participate in the second phase of the study, most did so (2,057 out of the original sample of 2,326 students). Participants were paid $40 for part one and $10 for part two. Table 1 describes the study sample with respect to several background characteristics and compares this group of students with the GRE General test-taking population for 1997-98. In general, the study sample is similar to the GRE population, although some differences are apparent. For example, our sample contains greater proportions of African American and Asian American students, and fewer White students, than does the GRE General Test population.

Instruments

Course-related writing samples. Students who wished to participate in the present study were asked to bring copies of two course-related writing samples to the testing center when they wrote their essays for Schaeffer et al.. To help them select acceptable writing samples, these students were told that examples could include essays, term papers, book reviews, lab reports, proposals, and case studies. They were also told that the samples should be 10 pages or less, written within the previous six months, and completed on a typewriter or word processor. They were instructed that one sample should be typical of their writing -- neither their best nor their worst -- and that the other should be somewhat lower in quality than the first. (The specification for a second, lower-quality sample was motivated by the 1996 finding by Powers et al. that, despite the request for a typical writing sample, most students submitted samples that had received very favorable evaluations by their professors; a majority reportedly received grades of A or A-).

GRE essays. The GRE Issue prompts administered in the first part of the study presented participants with a brief quotation that stated or implied an issue. These students were then asked to compose an essay in which they explored the implications of the issue, while developing their own perspective in a clear and logically compelling manner. Similarly, the Argument prompts presented a brief argument and asked examinees to write an essay in which they discussed their opinions of the validity of the argument. Study participants were given 45 minutes to compose the Issue essay and 30 minutes for the Argument essay.

Background questionnaire. The background questionnaire, also administered in the first part of the study, sought information from participants about such characteristics as their ethnicity, gender, citizenship, best language, undergraduate major field, degree objective, current educational status, and undergraduate grades (overall, in the major field, and in courses that required considerable writing). The students were also asked a number of questions about the testing experience and their reactions to it. One of the questions concerned the amount of effort that they had invested in writing each essay, compared with how much effort they would have invested if the test "had counted for admission to graduate school." Answer options included "about the same," "somewhat less," and "considerably less." Much of this information has been reported by Schaeffer et al. (in preparation).

Writing study questionnaire. The purpose of this questionnaire, administered only to those students participating in phase two of the study, was to gather information about the writing samples that study participants submitted. Participants received the questionnaire in advance of the test date and were

asked to complete it in advance and bring it with them to the testing center. It asked for descriptions of the assignments that elicited the writing samples they submitted, their perceptions of the cognitive requirements of the assignments, the grades they received on the samples, and a variety of other information about the samples, including several of the characteristics discussed by Norton (1990). Participants were also asked to compare their writing skills with those of their peers, to report information about any writing-related accomplishments (such as publications and work experiences), and to indicate the degree of success they had achieved with various types of writing and various aspects of the writing process.

As in the Powers et al. (1996) study, information about student accomplishments in writing was derived from responses to a brief inventory that was embedded in the writing study questionnaire. This inventory was an abbreviated version of a longer, more comprehensive measure based on the work of Baird (1979) and extended by Lawrence Stricker (Stricker & Rock, 1996). The particular adaptation used here involved only questions about writing-related accomplishments -- not accomplishments in other areas (such as science and music, for instance) -- and respondents were not asked to provide documentation of their achievements, as was required by Baird. The major categories of variables derived from the writing study questionnaire -- or what we have called nontest indicators of writing skill -- are summarized in Table 1.

Procedures

General design. For part one of the two-part project, participants were randomly assigned in equal numbers to one of four conditions, based on writing assignments:

1. two Issue essays
2. two Argument essays
3. one Issue essay followed by one Argument essay
4. one Argument essay followed by one Issue essay

Prior to taking the test, all participants received a tutorial on using a word processor to compose their essays. A total of 40 prompts -- 20 Issue and 20 Argument, all previously field tested -- were used in the study. Each prompt was administered with equal frequency, and equally often in the first and the second positions. Further details regarding the design are provided in the report by Schaeffer et al. (in preparation).

Evaluation of GRE essays. Both the Issue and the Argument essays were scored holistically according to six-point scales that were specific to each assignment and were thus slightly different. For Issue essays, a score of 6 (outstanding) was assigned to essays that presented a cogent, well-articulated analysis of the issue and demonstrated mastery of the elements of effective writing. The latter entailed:

- developing a position on the issue using insightful reasons and/or persuasive examples
- sustaining a well-focused, well-organized discussion
- expressing ideas clearly and precisely
- using language fluently, with varied sentence structure and effective vocabulary
- demonstrating superior facility with the conventions of standard written English.

At the other end of the Issue scale, the lowest score of 1 (fundamentally deficient) was assigned to papers that:

- provided little evidence of an ability to develop or organize a coherent response to the topic
- had severe and persistent errors in language and sentence structure
- contained a pervasive pattern of errors in grammar, usage, and mechanics that severely interfered with meaning

Argument essays were scored on a scale that also ranged from a high score of 6 to a low score of 1. In this case, a score of 6 (outstanding) was awarded to Argument essays that presented a cogent, well-articulated critique of the argument and demonstrated mastery of the elements of effective writing by:

- clearly identifying important features of the argument and analyzing them insightfully
- developing ideas cogently, organizing them logically, and connecting them with clear transitions
- effectively supporting the main points of the critique
- demonstrating control of language, including diction and syntactic variety
- demonstrating facility with the conventions of standard written English

At the other end of the score range, a 1 (fundamentally deficient) was assigned to essays that:

- provided little evidence of an ability to either understand and analyze the argument or to develop an organized response
- had severe and persistent errors in language and sentence structure
- contained a pervasive pattern of errors in grammar, usage, and mechanics that resulted in incoherence

For both the Issue and the Argument prompts, essays that were off-topic (or merely repeated the topic) were given a score of 0.

The essays were evaluated by specially trained college and university faculty from a variety of academic fields and with special interest and expertise in writing. Each essay was scored independently by two readers from a pool of 18. A third reader resolved any scoring discrepancies greater than one point, which occurred for about 2% of both the Issue and the Argument essays.

Schaeffer et al. (in preparation) have reported correlations of .62 between two Issue essays, .51 between two Argument essays, and .50, on average, between one Issue and one Argument essay. The reliability of an Issue/Argument composite was estimated to be .63 to .70. This estimate was based on a formula provided by Feldt and Brennan (1989; see formula 11, p. 112), in which the Argument and Issue essays were treated as separate test parts.

Evaluation of course-related writing samples. Scoring the student-submitted writing samples presented a special challenge for scorers in the Powers et al. (1996) study, in part because participants were given considerable latitude in the kinds of writing samples they could submit and were not asked for information about the precise nature of the assignments. As stated earlier, however, we intended to address this problem in the current study by systematically classifying all of the students' writing samples.

In the 1996 study, all writing samples were scored on a six-point holistic scale, according to how well the writer communicated his/her ideas. The scoring rubric was the same as that used to evaluate the GRE Issue essays. 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 their own background knowledge was insufficient for understanding a paper's content.

For the current study, we enlisted four experts in writing instruction and assessment to help us modify these scoring procedures. The first major step in the process was a two-day session at which the four experts, all from different institutions, met with the principal investigators to consider methods for scoring the course-related writing samples. In preparation for this session, we sorted the papers by assignment category and, for each category, selected a small subset of papers that, in our judgment, exemplified a range of quality. We also attached to each paper a copy of the student's writing study questionnaire, which included a description of the assignment. Then, the group read, scored, and discussed a sample of approximately 80 papers from across the various categories in order to determine whether to tailor a scoring guide for each category.

After this preliminary session, it became apparent that the assignment classifications that were provided by student participants tended to be idiosyncratic and did not match expert judgments. For this reason, the consultants advised us not to use these categories to evaluate the papers. Also, the group concurred that certain categories of papers (résumés, poems, and journal entries, for example) did not, by their very nature, exhibit the qualities of analytic writing that were of most interest to us, and should therefore be eliminated from scoring altogether.

It was also decided that the most effective method for evaluating the papers would be to merge the GRE scoring guides for the Issue and Argument essays and to expand the criteria slightly in order to focus on the complexity of thought rewarded in both of the guides. The resulting GRE Validity Scoring Guide reflected an expanded concept of critical thinking that was defined by one of the consultants (Art Young) as "scholarly habits of mind," an especially useful term in describing the focus of our evaluation. This guide employed the same six-point scale and labels (outstanding, strong, adequate, limited, seriously flawed, and fundamentally deficient) as the Issue and Argument guides, and the defining features at each level combined elements of both rubrics.

The second major step in evaluating the course-related writing samples was a two-day training session in which 21 college and university faculty (all teachers, and experienced evaluators, of writing) received training on how to evaluate the papers using the GRE Validity Scoring Guide. The session was conducted by a highly experienced scoring leader, who was also one of the four consultants we had enlisted earlier in the process. During this training session, all readers evaluated and discussed a range of papers from each assignment category. On the advice of our consultants, the readers were instructed to score each writing sample on its own merits, based on how well it met the criteria set forth in the Validity Scoring Guide -- not on what the original assignment might have been. In fact, readers were not given access to information about the nature of the assignment, the grade the assignment received, the length of time the student spent composing the assignment, or any other details from the writing study questionnaires.

Moreover, readers were also asked to ignore any teacher comments that could not be removed from the papers.

Near the end of the two-day training session, the readers began scoring the samples. For practical reasons, most of the samples were to be scored by first readers only. But first, to gauge the reliability of the readers, a subset of 80 samples was scored independently by second readers. The correlation between readers scoring the subset was relatively high ($r = .78$), and the discrepancy rate was reasonably low (8%, with "discrepancy" defined as a set of scores more than one point apart). Therefore, we decided to let readers evaluate a significant proportion of the papers at home, and agreed that a single score would be adequate for each paper. The scoring leader and Educational Testing Service (ETS®) researchers were available to the readers on an as needed basis, and this support proved especially useful when readers questioned whether or not a paper provided enough text to be evaluated.

Analyses. Analyses were designed to address three areas of interest:

- the relationship between performance on GRE essays and nontest indicators of writing skill
- the role of the characteristics of students' course-related writing samples as a moderator of the relationship between GRE essays and writing samples
- the possibility of differential validity of the writing assessment for subgroups of students

Correlational analyses were used to address both the relationship between performance on GRE essays and nontest indicators of writing skill and the role of differences among the characteristics of students' course-related writing samples. Hierarchical regression analysis (Cohen & Cohen, 1983) was used to investigate the possibility of differential validity of the writing assessment for subgroups of students. Specifically, the scores assigned to course-related writing samples were regressed on scores assigned to GRE essays. A variable indicating subgroup membership (that is, gender, ethnicity, language background, or undergraduate major field) was added next, and the extent to which this addition improved prediction was tested. Finally, a term reflecting the joint contribution, or interaction, of subgroup membership and GRE essay performance was added to the prediction equation, and the extent to which this variable increased the multiple R^2 was tested. This final test, an indication of the parallelism of regression lines, was used as the index of differential validity for subgroups. This analysis was repeated, using each of the other nontest indicators of writing skill in place of the scores assigned to course-related writing samples, and for each combination of GRE essays.

Results

Study Sample

As indicated earlier, the validity sample used for the second part of the research project was a slightly smaller subset ($N = 2,057$) of the sample ($N = 2,326$) used by Schaeffer et al. (in preparation). Like Schaeffer et al., we included only study participants who wrote at least 500 bytes of text

in response to each GRE essay prompt. The aim here was to eliminate participants whose motivation was questionable.[2]

Description of Writing Samples

We wanted to describe students' course-related writing samples in as much detail as possible in order to identify any features of academic writing that might render the samples more (or less) strongly related to performance on the GRE Writing Assessment. To this end, in the writing study questionnaire, we asked participants to characterize each of their writing samples in terms of the cognitive requirements of GRE essays. These requirements, along with other features of the samples, are shown in Table 3. We were also interested in whether any differences between the two samples of a given participant would account for differences in what our readers deemed to be the quality of those samples.

Slightly fewer than a third of the students reported that both of their course-related writing samples had been written for the same instructor. A majority of the samples (nearly 60%) were described by students as essays, research papers, or book reviews and analyses; some were placed in multiple categories by students. The vast majority was written outside of the classroom, with topics being chosen by students slightly more frequently than by their professors. A small minority of topics was determined jointly by students and professors. More than three fourths of the samples were written within the year prior to the study, and about one quarter within the previous month.

Rarely did students spend 10 hours or more composing their samples: About half of the papers were written in 3 to 9 hours, and nearly 40% were written in two hours or less. The average estimated length of samples was slightly more than 1,100 words, although length varied considerably: About a quarter of the samples were under 500 words, and about 12% exceeded 2,000 words. A substantial majority of participants said that they had received little or no help in drafting, editing, or revising their samples.

A significant majority of the samples had received grades of A+, A, or A- from professors. According to the students, the grades on about 66% of the papers had either some effect or a moderate effect on course grades. For nearly half of the samples, students felt that a substantial amount of the instructor's grade was based on the content or substance of the sample. Somewhat less often (for about a third of the samples), students reported that a substantial portion of their grades was based on the characteristics of their writing.

With respect to the cognitive requirements of the writing assignments, students reported that their samples more often required the recall of information, rather than its retrieval (from the library or the Internet, for example). Samples varied considerably with respect to whether or not they required the consideration of various perspectives and viewpoints, the construction of an argument, or the use of evidence to support a position (the skills elicited by GRE essays). According to respondents, about half the samples required little if any critique or analysis of arguments or positions.

[2] Hunter Breland (personal communication, Dec. 21, 1998) challenged this assumption, however, reasoning that instead of being unmotivated, participants who wrote very little may not have been able to think of much to write. He posits that the exclusion of these cases from our analyses may have decreased the correlations between GRE essay scores and the nontest indicators of writing skill.

It appears that we were successful in our attempt to avoid collecting only the students' very best writing samples. As we had hoped, our request for a second, lower-quality sample yielded, on average, two samples that did differ in quality. The lower-quality samples had, on average, received lower grades from professors, and in fact, they also received lower scores from the readers trained for our study (see Table 4).

To verify the consistency of participants' self-reports about the characteristics of their writing samples and other information, we recontacted a sample of 45 participants approximately one week after they had returned the original questionnaire, asking them to complete the questionnaire again. As an estimate of test-retest reliability, product-moment correlations were computed between responses from the first and second questionnaires. For individual questions that pertained to characteristics of the writing samples, the median correlation between the first and the second responses was .55 for one writing sample and .59 for the other. Interestingly, the students' perceptions of the cognitive requirements of the assignments that gave rise to their writing samples were the least consistently reported.

Relationships Among Criteria

Intercorrelations. Table 4 also provides correlations found among the nontest indicators of writing skill for which data were collected. The only relatively strong correlation (r = .63) was between self-reports of success with various steps of the writing process and success with different kinds of writing. Moderate correlations were noted among self-reports, self-comparison with peers, and grade point average in courses that required considerable writing (rs in the .40s). Correlations among all other indicators were weak.

Reliability. With respect to the reliability of the nontest indicators, we noted the following. For the sample of 45 students who completed two questionnaires, correlations between indices based on their first and second responses were .63 for the steps of the writing process and .71 for kinds of writing. The number of writing accomplishments reported on the first and second questionnaires correlated very strongly (r =.84). The first and the second reports of grade-related variables were relatively consistent also. For example, reports of grades received on the samples correlated .78 and .62 for the two samples, and reports of grades deserved correlated .74 and .69. Reports of the extent to which the samples contributed to course grades, and the basis on which professors graded the samples, correlated less strongly, with a median correlation of .53.

Table 5 contains greater detail about the correlations between three indices of the quality of the writing samples -- scores assigned by our readers, grades assigned previously by professors, and grades students believed they *should* have received on the samples. As Table 5 reveals, although students and professors exhibited relatively good agreement on grades, there was little agreement between readers' scores and evaluations by either students or their professors. Appraisals of the two writing samples bore only a moderate relationship to one another, regardless of the source: students (r = .50), professors (r = .42), or the readers trained for the study (r = .40). Thus, we were unable to predict to any substantial degree the quality of students' undergraduate course writing, even from another similar writing sample. This lack of generalizability in the criterion necessarily limits the extent to which it can be expected to relate to performance on the GRE Writing Assessment.

As Table 6 shows, the two writing samples tended to differ with respect to each of the characteristics for which we obtained information. For instance, the amount of information retrieved to write one sample did not correspond very strongly to the amount retrieved for the second sample (r = .29).

A student's samples tended to be most similar with respect to the bases on which instructors had awarded grades (r = .51 for content and r = .54 for mechanics) and when the samples were written (r = .54).

We attempted to determine the extent to which differences between the characteristics of the two samples may have lessened the generalizability of the writing sample criterion. To accomplish this, we computed the correlation between readers' scores on each of the two writing samples and the samples' similarity to one another, using each characteristic as a measure of that similarity. For example, did the two scores bear a stronger relationship with one another if the two samples were assigned by the same professor, or by different professors? Did the scores relate more strongly if the cognitive demands were similar for both samples?

Table 7 shows these correlations between readers' scores in relation to the similarity of various characteristics. In general, the scores did bear a stronger relationship to one another when the samples had similar characteristics. While this finding proved true for nearly all of the features of the samples, some characteristics appeared to be more instrumental than others in determining the correlation between the quality of samples: the amount of time devoted to writing the samples, the degree to which they contributed to course grades, and whether the samples were written in or outside class.

Table 8 shows how characteristics of the writing samples related to the scores assigned by readers. Note that the estimated length of the samples was most strongly related to the scores. Other characteristics were also related, including the amount of time students devoted to composing the samples and the role of the samples in determining course grades. Other features bore only very weak or inconsistent relations to readers' evaluations of the samples.

Relationships Between Performance on GRE Essays and Nontest Indicators

Overall. Table 9 shows correlations between scores on the essays proposed for the GRE Writing Assessment and the quality of students' writing samples (as scored both by trained readers and as graded by students' professors), as well as correlations between essay scores and each of several alternative indicators of writing skill. Correlations corrected for attenuation have also been provided. For writing samples, correlations between the two samples have been used as estimates of reliability.

The patterns of relationships between the scores and the various criteria were relatively similar for each of the three different combinations of two GRE essays administered to students (two Argument, two Issue, or one of each). Essay scores were slightly but consistently more strongly related to trained readers' evaluations of writing samples than to any of the other indicators. When adjusted for the weak correlation between readers' evaluations of sample quality, the correlations increased to about .50 for the Issue/Argument essay combination and nearly to this level for the other two combinations. Performance on GRE essays was only weakly related to the grades that professors had assigned to the samples. Grade point average in courses that required considerable writing was the next most highly related indicator; overall grade point average and grade point average in major field of study, on the other hand, were less strongly related. The various self-reports of success with writing and self-comparisons with peers were related only modestly to GRE essay scores. Reports of accomplishments in writing bore a negligible relationship to essay performance.

Table 10 and Figures 1-8 present the same data in another form, showing performance on each of the nontest criteria by GRE essay score level (for the one Argument, one Issue combination only). With the exception of accomplishments in writing, each of the indicators exhibited higher means at each successively

higher level of GRE essay performance. Moreover, in Cohen's (1977) effect size terms, there is a "large" mean difference on each indicator (again, with the exception of writing accomplishments) between the highest and the lowest GRE essay score levels. These differences range from approximately 0.8 to 1.9 sds.

By characteristics of writing samples. Table 11 provides correlations between scores for each combination of proposed GRE essays and readers' scores on the course-related writing samples, for samples having different characteristics. On inspection, the correlations do not readily suggest any features of the writing samples that may render them consistently more (or less) predictable from success on any combination of GRE essays. Our hypothesis about possible moderating features was, therefore, largely unsupported. For example, despite the presumably greater control over in-class essays, and their greater similarity to GRE essays, this kind of writing sample did not appear to relate consistently more strongly to success with the proposed GRE essays than did other, less similar writing samples. Nor, for example, did the proposed GRE Argument essays relate more strongly to writing samples that required close analysis of text than to those that did not.

Differential Subgroup Validity

To assess the possibility that the GRE Writing Assessment might have different meaning for subgroups of test takers, we conducted a hierarchical regression analysis by gender, ethnicity, best language, and undergraduate major. That is, for each of these classifications, we predicted performance on each of the nontest indicators in turn using GRE essay scores and subgroup membership. Next, we added to the regression equation a product variable reflecting the interaction of essay score and subgroup membership, computing the significance of the increment in R^2 due to the addition of the interaction term. Analyses were repeated for each of the three combinations of GRE essays.

Table 12 shows the results of these analyses for those 10 cases out of 72 that yielded a significant increase in R^2 due to the addition of the interaction term. (Table 13 summarizes the results of all 72 analyses by providing the increase in R^2 and its significance.) Six of the significant results were found for minority students, but even within this group, the results show no consistent pattern, as they are spread among each of the three combinations of GRE essays. Furthermore, the least predictable group differed from analysis to analysis. Tables 14-17 and Figures 9-16 reveal the relatively similar relationships observed for each subgroup classification. Tables 18-20 show the extent of overprediction, or underprediction, of three nontest indicators by gender, ethnicity, and best language when regression equations based on the total sample are used to make estimates. As can be seen, when it occurred, underprediction was slight. Moreover, it more often occurred for men, white examinees, and examinees whose best language was English.

Finally, we also examined whether estimates of validity may have differed according to the amount of effort that participants expended when they wrote their essays. A substantial proportion of examinees (44% on the first essay and 40% on the second) reported that they exerted the same amount of effort they would have exerted if the test had counted for graduate admissions, and an equally large portion expended only somewhat less effort (45% and 49% on the first and second essays, respectively). Relatively few students reported that they exerted considerably less effort (10% and 12 %, respectively). The interaction of effort with essay performance did not add significantly to prediction in any of the analyses, suggesting that the relationship between GRE essay performance and each nontest indicator did not vary according to the effort that participants put forth when writing their essays.

Discussion

Summary of Major Findings

The study described here had several purposes. The primary aim was to provide further evidence of the validity of GRE Writing Assessment scores. In particular, we were interested in the extent to which these GRE scores reflect the writing skills expected in academic settings. More specifically, our objective was to establish the relationship between performance on GRE essays and a variety of nontest indicators of student writing skills -- especially the quality of students' course-related writing samples. Analyses revealed that the various nontest indicators we examined bore modest and differential relationships to performance on GRE essays. The quality of course-related writing samples exhibited the strongest relationship to GRE scores, and self-reports of accomplishments in writing exhibited the weakest. The nontest indicators themselves were only modestly related to one another, suggesting that they probably tapped somewhat different aspects of the writing construct. Regardless of the particular combination of essays that students were asked to write (two Issue essays, two Argument essays, or one of each), performance on the GRE Writing Assessment was more highly related to the quality of course-related writing samples than to any other indicator of writing skill. We speculate that all of these relationships would have been stronger if test takers had had a greater opportunity to familiarize themselves with the writing tasks before the study -- as they would on an operational test -- thus reducing one irrelevant source of test difficulty.

Most of the evidence we gathered pertains to the *convergent* validity of the writing assessment -- that is, its relationships to variables with which it *should* correlate. We were able also, however, to provide modest evidence of the *discriminant* validity of the measure. In particular, GRE essay scores correlated in the expected manner with several aspects of undergraduate grades. The relationship was highest with grades in courses that required considerable writing, intermediate with grades in all courses, and lowest in major field courses, where mastery of subject matter presumably weighs more heavily in the awarding of grades than does writing skill.

The strength of the relationship between GRE essays and the two writing samples was moderate but noteworthy in light of the fact that judgments of the quality of the course-related writing samples were only moderately related. Thus, as might be suspected, criteria against which tests are validated -- in this case, course-related writing -- can suffer from the same lack of generalizability that has characterized performance assessments in general (see, for example, Linn, 1993; Shavelson, Baxter, & Pine, 1992; Shavelson, Baxter, Xiaohong, 1993) and writing assessments in particular (see Breland, Camp, Jones, Morris, & Rock, 1987, for instance).

In an effort to understand the lack of generalizability of our primary criterion -- the two course-related writing samples -- we conducted additional analyses. This further inspection revealed that the course samples differed with respect to a wide variety of characteristics, such as the nature of the writing assignments, the amount of time students devoted to writing them, and the amount of help that writers received when composing them. The amount of time students spent writing the samples, the role of the samples in determining course grades, and especially the length of the samples were all related to evaluations of the quality of the samples.

When the course-related writing samples were classified according to their characteristics, we were unable to detect any consistent patterns that suggested the scores on the GRE essays were more, or less, strongly related to some kinds of writing samples than to others. We hypothesized, for example, that performance on Argument essays would be more strongly related to writing samples that required close

analysis of text than to those that did not. However, no such relationship was found. Thus, from this viewpoint, our results do not reveal any particular discriminant validity of the GRE Writing Assessment. Neither, however, do they reveal any particular lack of generality over several classifications of course samples, as we were unable to identify any characteristics of the samples that rendered their quality any more or less predictable from GRE essays.

The various combinations of GRE essay prompts that were assigned to students each seemed to bear much the same pattern of relationships with the nontest criteria. Neither the combination of two Issue essays nor the combination of two Argument essays proved to relate more strongly to the nontest indicators than did the combination of one Issue essay and one Argument essay. And, no consistent patterns were detected that might suggest that the relationship between GRE essay scores and the various nontest criteria differed for any of the subgroups of examinees considered in the study.

Limitations of the Study

Each of the nontest indicators of writing skill included in the study contributed in at least some small way, we believe, to the overall appraisal of writing skill. Arguably, course-related writing samples were the most compelling of the indicators studied here, but -- like all of the nontest indicators -- they have their own characteristic strengths and limitations as validation criteria. One concern with course-related samples, for instance, is their lack of standardization: The samples we collected were written for a variety of different purposes and audiences.[3] As mentioned earlier, although participants in the Powers et al. (1996) study were asked to submit *typical* samples of their writing, they seemed to furnish instead the *best* examples. For the current study, by requesting a second sample that was not as good as a typical sample, we were modestly successful in getting a greater number of samples that were, if not more typical, at least less exemplary than they might have been. And, although the samples still constituted a heterogeneous collection, in the current study we were able to describe the variation among samples and consider it in our analyses. In short, we attained a substantially better understanding of the primary criterion in the current study than we had in the earlier one. To the best of our knowledge, our effort here represents one of the very few times that holistic scoring has been applied on a large scale to such disparate papers from such a variety of academic fields of study.

As in the Powers et al. (1996) study, we also gathered student opinions about their own writing for the present study. Comparisons with peers and other self-assessments have been shown to have significant validity in a number of different contexts (as gauged by correlations with other objective indicators of ability; see Mabe and West, 1982). This is true particularly in academic settings (see, for example, Baird, 1976), and most especially in self-assessments of language skills (Blanche and 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 their experiences. Self-evaluations suffer, however, from their potential subjectivity: While individuals may sometimes be the best judges of their own capabilities, they may also sometimes be the worst.

We also collected self-reports involving more factual matters (grades and accomplishments, for example), but these may be suspect as well. Several factors can cause contamination of self-reported data: forgetting (or selective remembering), responding according to socially accepted norms (especially for some kinds of information), or deliberate falsification. Nonetheless, self-reports have proven to have sufficient

[3] Breland (personal communication, Dec. 21, 1998) has suggested that instead of conducting an overall analysis based on the total sample, we might have examined correlations within each of the 26 institutions at which data were collected and then averaged the results over all institutions. This strategy might have produced somewhat higher correlations.

accuracy for a wide variety of purposes. This applies in particular to self-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).

In summary, we make no claim that, as a collection, the nontest indicators of writing investigated here constitute a comprehensive theoretical framework that spans the construct of writing proficiency. However, we have employed several methods -- such as self-assessments, self-reports, and independent evaluations -- to measure criterion performances, and we have included measures that seem to us to reflect different aspects of writing skill. The various indicators differed also with respect to the time periods that they represent, and they vary with regard to whether the process or the product of writing was the focus. The indicators also differed according to the evaluator: Students, their professors, and the readers trained specifically for this study each brought unique perspectives to bear. Our hope is that, in the aggregate, we have converged on a meaningful criterion of student writing skill.

Conclusions

In summary, we offer the following conclusions:

- From the perspective of this study, administering a combination of one Issue essay and one Argument essay is at least as valid as administering either two Issue or two Argument essays.

- The quality of writing performed in an academic setting varies appreciably from one occasion to the next. Therefore, it is difficult to predict performance on one assignment from performance on a similar one with a high degree of accuracy.

- Performance on the GRE Writing Assessment relates reasonably well to the quality of course-related writing samples, given the unpredictability of individual samples. A more comprehensive criterion -- involving multiple course-related samples -- would be necessary to estimate the predictive power of the Writing Assessment more accurately.

- Performance on the GRE Writing Assessment relates modestly to a number of other, nontest indicators of student writing skills. It does not, however, seem to relate to self-reports of high-level accomplishments in writing.

- Performance on the GRE Writing Assessment relates in the expected manner to subsets of undergraduate grades, correlating more strongly with grades in courses that require considerable writing than with grades overall or with grades in the major field of study.

- According to the way in which we have characterized course-related writing samples for this study, performance on the GRE Writing Assessment does not appear to relate differentially to any particular kinds of writing samples, nor does the relationship seem to differ consistently according to the circumstances under which the samples were written.

- Scores derived from the GRE Writing Assessment appear to have similar meaning for a variety of subgroups of test takers, as reflected in comparable associations between performance on the writing assessment and each of several nontest indicators of writing skill for several subgroups.

Implications of the Study

At the most general level, the major outcome of the effort described here is the collection of additional information about the two writing tasks -- "Present Your Perspective on an Issue" and "Analyze an Argument" -- that will constitute the GRE Writing Assessment when it is introduced in 1999. As Cumming (in press) and others have pointed out, critics of writing assessments have questioned the correspondence between writing assessment tasks and writing that is required in academic settings. The relationship is regarded as especially precarious when academic writing is accomplished over a significant period of time with ample resources on which to rely.

Powers and Fowles (1998) have also cited critics' concerns about whether impromptu writing performed within relatively constrained time limits can adequately reflect the kinds of writing that are performed in academic contexts. As one faculty member stated:

> Most graduate student writing tasks do not consist of being given an unfamiliar topic and being asked to write about it off the top of the student's head. Most writing tasks consist of doing research to discover a set of facts and then pulling these facts together clearly, accurately, and in interesting ways (p. 140).

This lament is symptomatic of the fact that, for practical reasons, most assessments are never able to fully represent every important aspect of the theoretical construct of interest. For example, most standardized writing assessments are not, in the limited amount of time that is typically available, able to capture aspects of writing -- such as careful planning and revision -- that are regarded as integral facets of the writing process (see, for example, Hayes, 1996). Thus, there is only an imperfect correspondence between the construct of academic writing and how it is typically measured. Nonetheless, our results suggest that there is a meaningful relationship between the writing skill manifested in an academic setting as students attempt to display the knowledge they have acquired (see Hale, Taylor, Bridgeman, Carson, Kroll, & Kantor, 1996, for example) and how this skill is quantified by the GRE Writing Assessment.

Perhaps one of the most useful, preoperational outcomes of the present study is information about the extent to which test score inferences are equally meaningful -- and in this sense, fair -- to several classifications of test takers. There was little if any evidence to suggest that the GRE Writing Assessment will be any more or less valid for some groups of test takers than others.

The results also provide, for each of the two proposed GRE writing exercises, a firmer basis for understanding the meaning of scores based on them, as required by professional standards for educational and psychological testing. We sought to establish the kinds of writing (and the types of circumstances under which that writing is performed) that are most, and least, highly related to performance on GRE Issue and Argument tasks. Our objective was to learn more about the conditions under which inferences based on GRE essay scores can be made with some degree of confidence, as well as some of the circumstances that may necessitate more cautious claims. Establishing these boundaries is, as Messick (1989) and others have noted, a hallmark of test validation. We have provided, in Messick's terms, evidence of the generality of the interpretation of scores across groups and settings or contexts, showing how stable and circumscribed their meaning is likely to be. The extent to which a measure displays the same pattern of relationships for different population groups and under different conditions is, according to Messick, a pervasive and perennial validity question. We have, we believe, provided important evidence of this stability, as it pertains to the GRE Writing Assessment.

We note, as a postscript, that when asked if they would be willing to participate in follow-up research on the GRE Writing Assessment, fully 94% of the students responded positively. Furthermore, nearly all of these participants provided the name, address, and telephone number of a parent or relative who would know how to contact them in the future. Thus, we are positioned to conduct further research on the relationship between performance on the GRE Writing Assessment and success with writing at the graduate level, should the GRE Board be inclined to pursue this kind of research. In addition, any further research should attempt to collect GRE General Test scores in order to establish the relationship between scores on the GRE Writing Assessment and performance on each of the three components of the General Test. With these data, the independent, and incremental, contribution of the Writing Assessment to the GRE test battery can be determined.

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 Research Report No. 83-32). New York: College Entrance Examination Board.

Breland, H. M., Bridgeman, B., & Fowles, M. (in press). Writing assessment in admissions to higher education: Review and framework (Draft GRE Board Report No. 96-12). 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., Kubota, M. Y., & Bonner, M. W. (in press). The performance assessment study in writing: Analysis of the SAT II Writing Test. (ETS Research Report No. 98-X). Princeton, NJ: Educational Testing Service.

Cohen, J. (1977). Statistical power analysis for the behavioral sciences (revised edition). New York: Academic Press.

Cohen, J., & Cohen, P. (1983). Applied multiple regression/correlation analysis for the behavioral sciences (2nd ed.). Hillsdale, NJ: Erlbaum.

Cronbach, L. J. (1990). Essentials of psychological testing (5th ed.) New York: Harper Collins Publishers.

Cumming, A. (in press). The testing of writing in second languages. In C. Clapham (Series Ed.) & D. Corson (Vol. Ed.), Language testing and assessment. Boston: Kluwer Academic Publishers.

Feldt, L. S., & Brennan, R. L. (1989). Reliability. In R. L. Linn (Ed.), Educational measurement (3rd ed., pp. 105-146). New York: American Council on Education/Macmillan Publishing Company.

Freeberg, N. E. (1988). Analysis of the revised student descriptive questionnaire: Phase I (College Board Report No. 88-5; ETS Research Report 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 Research Report No. 89-49). New York: College Entrance Examination Board.

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

Hale, G., Taylor, C., Bridgeman, B., Carson, J., Kroll, B., & Kantor, R. (1996). A study of writing tasks assigned in academic degree programs (TOEFL Report No. 54; ETS Research Report No. 95-44). Princeton, NJ: Educational Testing Service.

Hambleton, R. K., & Rogers, H. J. (1991). Advances in criterion-referenced measurement. In R. K. Hambleton & J. N. Zaal (Eds.), Advances in educational and psychological testing: Theory and applications (pp. 3-34). Boston, MA: Kluwer Academic Publishers.

Hayes, J. R. (1996). A new framework for understanding cognition and affect in writing. In C. M. Levy & S. Ransdell (Eds.), The science of writing: Theories, methods, individual differences, and applications (pp. 1-27). Mahwah, NJ: Erlbaum.

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

Huot, B. (1990). The literature of direct writing assessment: Major concerns and prevailing trends. Review of Educational Research, 60, 237-263.

Laing, J., Sawyer, R., & Noble, J. (1987). Accuracy of self-reported activities and accomplishments of college-bound students (ACT Research Report Series No. 87-6). Iowa City, IA: The American College Testing Program.

Linn, R. L. (1993). Educational assessment: Expanded expectations and challenges. Educational Evaluation and Policy Analysis, 15, 1-16.

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.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), Educational measurement (3rd ed., pp. 13-103). New York: American Council on Education/Macmillan Publishing Company.

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

Norton, L. S. (1990). Essay-writing: What really counts? Higher Education, 20, 411-442.

Powers, D. E., & Fowles, M. E. (1996). Effects of applying different time limits to a proposed GRE writing test. Journal of Educational Measurement, 33, 433-452.

Powers, D. E., & Fowles, M. E. (1998). Effects of preexamination disclosure of essay topics. Applied Measurement in Education, 11, 139-157.

Powers, D. E., Fowles, M.E., & Boyles, K. (1996). Validating a writing test for use in graduate admissions (GRE Board Report No. 93-26b; ETS Research Report No. 96-27). Princeton, NJ: Educational Testing Service.

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.

Sawyer, R., Laing, J., Houston, M. (1988). Accuracy of self-reported high school courses and grades of college-bound students (ACT Research Report Series No. 88-1). Iowa City, IA: The American College Testing Program.

Schaeffer, G., Briel, J., & Fowles, M. (in press). Psychometric evaluation of the new GRE writing assessment. (Draft GRE Board Report No. 96-11). Princeton, NJ: Educational Testing Service.

Shavelson, R. J., Baxter, G. P., & Pine, J. (1992). Performance assessments: Political rhetoric and measurement reality. Educational Researcher, 21, 22-27.

Shavelson, R. J., Baxter, G. P., & Xiaohong, G. (1993). Sampling variability of performance assessments. Journal of Educational Measurement, 30, 215-232.

Stricker, L.J., & Rock, D.A. (1996). Measuring accomplishments: Pseudoipsativity, quantity vs. quality, and dimensionality (ETS Research Report No. 96-8). Princeton, NJ: Educational Testing Service.

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: Teachers of English as a Second Language (TESOL).

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

TABLE 1

Description of Study Sample versus GRE Test-Taking Population

Background Characteristics	Study Sample (N = 2057)	1997-98 GRE Population (N = ≈ 320,000)
Female (%)	61	59
Ethnicity (%)		
American Indian	<1	<1
African American	21	8
Mexican American	2	2
Asian American	18	5
Puerto Rican	1	1
Other Hispanic	4	2
White	50	78
Other	4	3
U. S. Citizen (%)	81	80
Best Language English (%)	83	94
Undergraduate Major Field (%)		
Natural Science	29	31
Engineering	15	12
Social Science	19	23
Humanities	11	14
Education	7	8
Business	9	3
Other	11	9
Graduate Degree Objective (%)		
Doctorate or higher	36	39
Less than doctorate	53	55
No plans	11	6
Current Educational Status (%)		
Junior or lower	28	n.a.
Senior	46	n.a.
Graduate school	24	n.a.
Other	2	n.a.

TABLE 2

Description of Other Nontest Indicators of Writing Skills

Category	Description
Comparison with peers	"How do you think your writing compares with that of other students in your field?" (Responses on a five-point scale from "well below average" to "well above average.")
Success with the writing processes	Reported success in college courses with four different writing activities: thinking about assignments, organizing, revising, and developing an effective writing style. (Responses were on a five-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 five-point scale ranging from "not at all successful" to "extremely successful.")
Grade average in courses involving writing	Grade point average in courses that required "considerable" writing.
Documentable accomplishments in writing	The number of writing-related accomplishments reported. (Responses were to an inventory asking about 10 different accomplishments such as writing and publishing articles, technical manuals, speeches, and so on.)

TABLE 3

Characteristics of Student-Submitted Course-Related Writing Samples

Characteristic	Sample X (Better sample)	Sample Y (Worse sample)
Assigned by same instructor (% yes)	31	31
Description (%)		
Essay	27	25
Research paper	19	15
Review or an analyses of book, etc.	14	16
Other	9	11
Summary or literature review	7	7
Report of an experiment	6	7
Case study	5	5
Plan/proposal	5	5
Multiple classifications	9	9
Where written (%)		
Entirely or mostly in class	5	7
Entirely or mostly outside class	95	93
When written (%)		
Within past month	26	27
One to twelve months ago	54	49
More than one year ago	20	24
Time spent composing (%)		
Two hours or less	31	45
About 3 to 9 hours	54	45
About 10 hours or more	15	10
Amount of information retrieved (%)		
Little if any	40	49
Some	20	20
A moderate amount	19	17
A substantial amount	21	15
Amount of information recalled (%)		
Little if any	27	29
Some	27	25
A moderate amount	21	19
A substantial amount	25	26

Table continued on next page

TABLE 3 (continued)

Characteristic	Sample X (Better sample)	Sample Y (Worse sample)
Sample required consideration of various perspectives/points of view (%)		
Little if any	27	32
Some	27	28
A moderate amount	24	23
A substantial amount	22	17
Sample required construction of an argument or evidence to support a position (%)		
Little if any	20	25
Some	19	24
A moderate amount	26	25
A substantial amount	35	27
Sample required critiquing/analyzing an argument/ position (%)		
Little if any	47	53
Some	20	19
A moderate amount	15	14
A substantial amount	18	14
Grade received (%)		
A+	22	15
A (or A-)	45	38
B+	11	11
B (or B-)	6	11
C+	1	3
C (or less)	1	2
Ungraded	14	16
Grade deserved (%)		
A+	26	17
A (or A-)	56	47
B+	12	20
B (or B-)	6	13
C+	<1	2
C (or less)	<1	1
Effects of sample on course grade (%)		
Little if at all	13	15
Some	33	39
A moderate amount	36	32
A substantial amount	19	14

Table continued on next page

TABLE 3 (continued)

Characteristic	Sample X (Better sample)	Sample Y (Worse sample)
Feedback received (%)		
Little if at all	30	37
Some	33	33
A moderate amount	25	20
A substantial amount	13	10
Grade based on content or substance? (%)		
Little if at all	7	10
Some	14	18
A moderate amount	30	32
A substantial amount	49	40
Grade based on characteristics of writing? (%)		
Little if at all	10	13
Some	24	27
A moderate amount	32	31
A substantial amount	34	29
Topic chosen by: (%)		
Student	49	45
Instructor	38	42
Jointly	13	13
Amount of help with drafting (%)		
Little or none	84	87
Moderate or substantial	16	13
Amount of help with editing/revising (%)		
Little or none	74	80
Moderate or substantial	26	20
Estimated length of sample (%)		
Less than 500 words	20	28
500 – 1,000 words	31	33
1,000 – 2,000 words	33	29
More than 2,000 words	16	10
Mean words	1239	1030
SD	898	753

Note. A total of 2,029 writing sample X and 2,013 writing sample Y were available.

TABLE 4

Correlations Among Several Nontest Indicators of Writing Skill

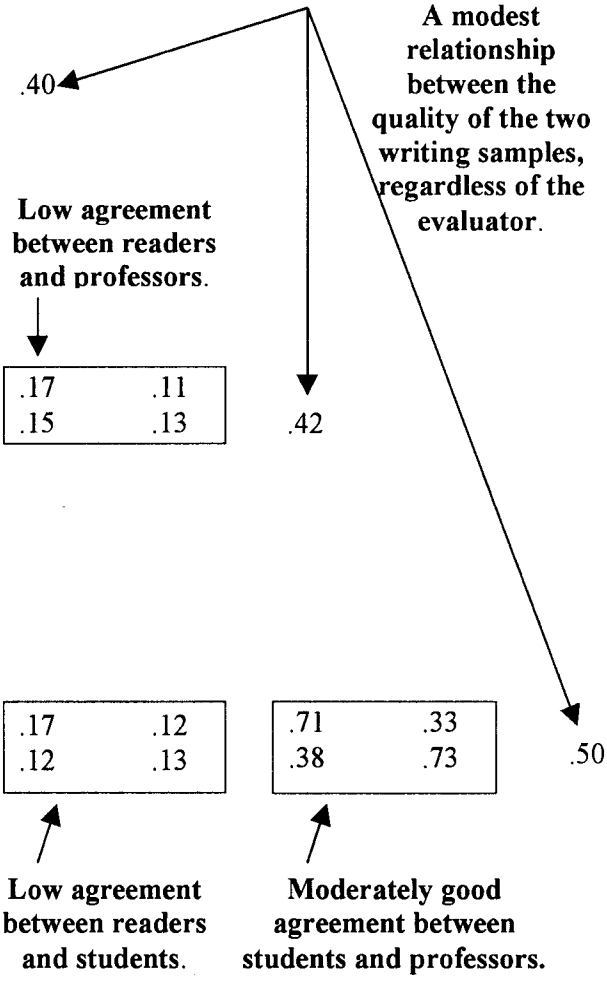
Indicator	M	SD	1	2	3	4	5	6a	6b	7a
1. Self-comparison with peers in major field	3.8	0.8								
2. Success with writing processes	3.9	0.6	.46							
3. Success with kinds of writing	3.8	0.5	.47	.63						
4. Grade average in writing courses	5.7	1.1	.41	.42	.43					
5. Accomplishments in writing	1.9	1.8	.24	.22	.24	.19				
6. Writing sample (professor-graded):										
a. sample X	4.9	0.9	.22	.27	.27	.35	.11			
b. sample Y	4.5	1.2	.25	.30	.25	.33	.05	.42		
7. Writing sample (reader-graded):										
a. sample X	3.8	1.2	.20	.12	.15	.24	.06	.15	.13	
b. sample Y	3.5	1.1	.15	.12	.14	.20	.03	.14	.17	.40

Note. Ns range from 1918 to 2039. Correlations of approximately .06 are significant at the .01 level, two-tailed. Professor grades are coded such that A+ = 6, A or A- = 5, B+ = 4, B or B- = 3, C+ = 2, and C or less = 1. Reader grades are on a 1 to 6 scale.

TABLE 5

Correlations Among Reader-, Professor- and Student-Assigned Grades on Course-Related Writing Samples

	M	SD	1(a)	1(b)	2(a)	2(b)	3(a)
1. Reader grade on:							
(a) sample X	3.8	1.2					
(b) sample Y	3.5	1.1	.40				
2. Professor grade on:							
(a) sample X	4.9	0.9	.17	.11			
(b) sample Y	4.5	1.2	.15	.13	.42		
3. Student assessment of:							
(a) sample X	5.0	0.8	.17	.12	.71	.33	
(b) sample Y	4.6	1.0	.12	.13	.38	.73	.50



Note. Readers made their evaluations on a 1 - 6 scale. Professors and students used a 1 - 6 scale, ranging from 1 = C or less to 6 = A+. N = 1933 for sample X and 1918 for sample Y.

TABLE 6

Correlations Between the Characteristics of Two Course-Related Writing Samples

Characteristic of Sample	r
When written	.54
Degree to which instructor(s) considered writing characteristics	.54
Degree to which instructor(s) considered content/substance	.51
Role in determining course grade	.45
Amount of help with drafting	.44
Amount of feedback received	.41
Amount of time devoted to composing	.40
Amount of help with editing/revising	.38
Where written (in class or outside class)	.36
Length of sample	.36
Degree of requirement to construct an argument or provide evidence	.35
Amount of argument critique/analysis required	.32
Person choosing topic (instructor or student)	.31
Amount of information recalled	.30
Amount of information retrieved	.29
Amount of consideration of various perspectives	.27

Note. Correlations are based on ns ranging from 1800 to 1997. All are statistically significant at the .01 level, two-tailed.

TABLE 7

Correlations Between the Quality of Two Writing Samples by Differences in Characteristics of the Samples

Comparison of Samples	Yes	No	n for Groups: Yes	n for Groups: No
Same amount of time devoted to both?	.50	.26	1,120	822
Same amount of information retrieved for both?	.49	.33	859	1,079
Same amount of information recalled for both?	.49	.34	816	1,117
Same amount of consideration of various perspectives required for both?	.49	.34	799	1,134
Both determined course grade to the same degree?	.49	.29	1,101	834
Both same kind of sample?	.48	.35	723	1,210
Same degree of requirement to construct an argument or provide evidence for both?	.47	.35	862	1,071
Same amount of feedback received for both?	.46	.34	998	935
Instructor(s) consider writing characteristics to the same degree for both?	.46	.33	1,154	779
Person who chose topic (instructor or student) the same for both?	.46	.33	1,096	845
Same amount of argument critique/analyses required for both?	.45	.36	969	964
Both assigned by same instructor?	.44	.38	585	1,311
Instructor(s) consider content/substance to the same degree for both?	.44	.35	1,205	728
Amount of help with editing/revising the same for both?	.43	.31	1,529	404
Both written in same time period?	.42	.37	1,277	656
Both written in same place (either in class or outside class)	.41	.24	1,785	148
Amount of help with drafting the same for both?	.41	.36	1,653	280

Note. With sample sizes of 1,500 and 300 and correlations in the range of .35 to .45, differences in r of about .10 are significant at the .01 level. Differences of about .08 are significant at the .05 level.

TABLE 8

Correlations Between Judged Reader Ratings of Writing Samples and Characteristics of Samples

Characteristic	Sample X	Sample Y
Length of sample	.51	.51
Time devoted to composing	.34	.31
Role in determining course grade	.24	.26
Amount of information retrieved	.22	.17
Degree to which instructor considered content/substance	.21	.20
Degree of requirement to construct an argument	.21	.19
Amount of feedback received	.12	.12
Degree to which instructor(s) considered writing characteristics	.12	.08
Where written (in class or outside class)	.08	.12
When written	.07	.13
Amount of consideration of various perspectives	.07	.10
Amount of argument critique/analysis required	.07	.07
Amount of help with editing/revising	-.01	-.01
Person chosing topic (instructor or student)	-.02	-.02
Amount of help with drafting	-.06	-.04
Amount of information recalled	-.07	-.07

Note. Correlations are based on ns ranging from 1705 to 1894. Correlations of approximately .06 are significant at the .01 level, two-tailed.

TABLE 9

Correlations Between Performance on GRE Essays and Other Indicators of Writing Skill

Criterion	Estimated Reliability of Criterion	1 Issue and 1 Argument Essay (n = 847)	2 Issue Essays (n = 458)	2 Argument Essays (n = 428)
			Writing Samples	
Writing samples (reader-graded):				
Both samples	.57	.38 (.50)	.37 (.49)	.35 (.46)
Sample X	.40	.33 (.52)	.32 (.51)	.27 (.43)
Sample Y	.40	.31 (.49)	.30 (.47)	.33 (.52)
Writing sample (professor-graded):				
Both samples	.59	.19 (.25)	.15 (.20)	.17 (.22)
Sample X	.42	.16 (.25)	.13 (.20)	.06 (.09)
Sample Y	.42	.16 (.25)	.11 (.22)	.16 (.25)
			Grades	
Grade point average:				
In writing courses	n.a.	.34	.28	.27
Overall	n.a.	.20	.21	.16
In major field	n.a.	.11	.10	.17
			Other Indicators	
Self-comparison with peers	.19	.29 (.66)	.22 (.50)	.16 (.37)
Success with various kinds of writing	.71	.26 (.31)	.27 (.32)	.19 (.23)
Success with writing activities	.63	.20 (.25)	.19 (.24)	.14 (.18)
Accomplishments (log)	.84	.07 (.08)	.11 (.12)	.02 (.02)

Note. For N = 400, correlations of .13 are significant at the .01 level, two-tailed. For N = 800, correlations of .09 are significant at the .01 level. Figures in parentheses are correlations that have been corrected for the unreliability of the nontest indicators.

TABLE 10

Performance on GRE Essays (One Issue plus One Argument) versus Performance on Nontest Indicators

Indicator		GRE Essay Score Level*					
		1.5*	2.5	3.25	4.0	4.75	5.5
	n	40	133	280	257	140	38
Sample X	M	3.3	3.4	3.7	4.0	4.4	4.7
	sd	1.0	1.3	1.1	1.0	1.0	1.0
	% 5 or 6	10	23	22	31	42	69
Sample Y	M	2.7	3.2	3.5	3.7	4.0	4.2
	sd	1.1	1.0	1.0	1.0	1.0	1.1
	% 5 or 6	5	8	16	23	27	37
GPA in writing courses	M	4.9	5.3	5.5	5.9	6.1	6.6
	sd	1.0	1.1	1.1	1.0	1.0	0.7
	% A+	5	15	21	30	42	67
GPA overall	M	4.7	5.1	5.2	5.4	5.6	6.1
	sd	1.4	1.2	1.2	1.2	1.1	0.9
	% A+	9	15	16	21	22	41
GPA in major	M	5.2	5.6	5.6	5.7	5.9	6.2
	sd	1.3	1.2	1.1	1.2	1.1	1.1
	% A+	21	28	24	31	32	50
Self-comparison with peers	M	3.2	3.6	3.7	3.9	4.1	4.2
	sd	0.7	0.8	0.8	0.7	0.7	0.7
	% above average	30	58	64	76	81	86
Success with writing process	M	3.7	3.8	3.9	4.0	4.1	4.3
	sd	0.6	0.6	0.6	0.6	0.6	0.4
Success with various kinds of writing	M	3.6	3.7	3.8	3.9	4.0	4.2
	sd	0.6	0.6	0.5	0.5	0.5	0.4
Accomplishments in writing	M	2.0	2.2	1.6	1.8	2.3	2.5
	sd	2.4	2.0	1.6	1.6	1.9	1.9

*Score levels represent the midpoint of score ranges. Scores assigned by four readers (two per essay) were summed, ranges established, and midpoints specified. For example, the level "1.5" represents the 4 – 8 total score range (i.e., 6 ÷ 4).

TABLE 11

Correlations of GRE Essays with Quality of Writing Samples by Characteristics of Samples

Characteristic of Writing Sample	1 Issue and 1 Argument Essay		2 Issue Essays		2 Argument Essays	
	Sample X	Sample Y	Sample X	Sample Y	Sample X	Sample Y
Kind of sample:						
Essay	.42	.43	.40	.21	.24	.32
Research paper	.27	.32	.21	.34	.24	.41
Report of experiment	.27	.21	.39	.40	.54	.04
Case study	.21	.30	.33	-.02	.37	.29
Literature review	.17	.29	.41	.34	.28	.28
Book review	.28	.17	.28	.45	.43	.26
Proposal	.37	.25	.14	-.34	.66	.34
Other	.24	.20	.51	.30	.21	.37
Multiple	.29	.39	.29	.34	.02	.43
Location:						
In class	.47	.25	.47	.31	.07	.29
Outside class	.30	.30	.31	.28	.26	.33
When written:						
Written past month	.34	.31	.37	.29	.21	.50
1-12 months ago	.28	.32	.27	.27	.23	.22
More than 1 year ago	.36	.29	.37	.31	.37	.30
Time devoted to writing:						
Two hours or less	.35	.31	.30	.36	.30	.29
3-9 hours	.31	.30	.32	.22	.27	.36
10 or more hours	.30	.36	.42	.27	.14	.43
Amount of information retrieved:						
Little	.34	.31	.42	.35	.29	.34
Some	.36	.31	.31	.33	.36	.44
Moderate	.45	.36	.32	.23	.23	.28
Substantial	.20	.35	.21	.23	.19	.33
Amount of information recalled:						
Little	.22	.29	.34	.36	.28	.30
Some	.39	.34	.31	.21	.23	.36
Moderate	.32	.24	.35	.30	.18	.32
Substantial	.34	.34	.27	.24	.37	.32

Table continued on next page

TABLE 11 (continued)

Characteristic of Writing Sample	1 Issue and 1 Argument Essay		2 Issue Essays		2 Argument Essays	
	Sample X	Sample Y	Sample X	Sample Y	Sample X	Sample Y
Amount of consideration of various perspectives:						
Little	.32	.18	.38	.37	.25	.27
Some	.36	.23	.22	.17	.23	.33
Moderate	.30	.40	.41	.27	.18	.36
Substantial	.27	.43	.25	.34	.37	.37
Amount of argument or evidence required:						
Little	.22	.26	.31	.17	.14	.28
Some	.35	.24	.18	.13	.23	.40
Moderate	.33	.30	.42	.45	.26	.24
Substantial	.29	.33	.23	.27	.23	.32
Amount of critique/analysis of an argument/position:						
Little	.32	.25	.33	.34	.37	.28
Some	.42	.33	.35	.11	.05	.39
Moderate	.24	.52	.25	.35	.12	.42
Substantial	.29	.24	.29	.32	.31	.30
Grade received:						
A+	.21	.41	.29	.37	.22	.26
A (or A-)	.32	.30	.32	.27	.24	.31
B+	.33	.40	.32	.33	.07	.27
B (or B-)	.37	.29	.18	.25	.10	.07
C+	—	—	—	—	—	—
C or less	—	—	—	—	—	—
Effect on course grade:						
Little	.24	.23	.43	.27	.43	.28
Some	.26	.24	.26	.27	.28	.38
Moderate	.39	.34	.35	.29	.22	.32
Substantial	.29	.34	.26	.37	.26	.42
Grade based on content/ substance						
Little	.39	.12	.28	.34	.36	.53
Some	.47	.34	.36	.30	.25	.35
Moderate	.25	.29	.28	.41	.22	.26
Substantial	.28	.31	.26	.14	.24	.30

Table continued on next page

TABLE 11 (continued)

Characteristic of Writing Sample	1 Issue and 1 Argument Essay		2 Issue Essays		2 Argument Essays	
	Sample X	Sample Y	Sample X	Sample Y	Sample X	Sample Y
Grade based on characteristics of writing:						
Little	.29	.21	.33	.30	.17	.31
Some	.35	.25	.40	.35	.38	.39
Moderate	.27	.39	.27	.20	.36	.37
Substantial	.38	.34	.31	.37	.13	.28
Topic chosen by:						
Student	.31	.31	.34	.35	.27	.42
Instructor	.34	.31	.24	.18	.20	.26
Jointly	.35	.27	.46	.31	.36	.17
Amount of help drafting:						
Little or none	.29	.29	.30	.26	.25	.30
Moderate or substantial	.39	.29	.42	.32	.31	.45
Amount of help editing:						
Little or none	.31	.31	.34	.26	.26	.30
Moderate or substantial	.30	.26	.36	.43	.25	.32
Estimated length:						
500 words or less	.29	.25	.26	.33	.39	.28
501 – 1,000 words	.27	.21	.14	.16	.25	.22
More than 1,000 words	.23	.25	.28	.13	.11	.26
Undergraduate major of writer:						
Natural Sciences	.28	.31	.18	.34	.18	.32
Engineering	.37	.30	.46	.38	.48	.36
Social Sciences	.34	.34	.38	.19	.24	.20
Humanities	.29	.17	.19	.34	.27	.43
Education	.41	.22	-.02	.01	.40	.13
Business	.32	.43	.26	.11	.20	.33
Other	.27	.29	.61	.20	.22	.33

Note. Ns for the Argument/Issue, two Issue, and two Argument combinations were 847, 458, and 428, respectively. Ns on which correlations are based for each of the three combinations ranged from a low of 41, 17, and 19 for essays written in class to 791, 434, and 400 for essays written outside of class. Ns for specific categories can be estimated from Table 3.

TABLE 12

Hierarchical Regression Statistics for Analyses Revealing a Significant Interaction Effect

Group/Indicator	GRE Essay Combination	Independent Variables	Cumulative R^2	Increase in R^2	F	df
Gender						
GPA in writing courses	2 Issue essays	1. Essay score (ES)	.07	.07	30.3***	1, 425
		2. ES, gender	.08	.02	8.2**	1, 424
		3. ES, gender, ES x gender	.10	.01	5.2*	1, 423
Success with various kinds of writing	2 Issue essays	1. Essay score (ES)	.08	.08	38.6***	1, 425
		2. ES, gender	.09	.01	2.9	1, 424
		3. ES, gender, ES x gender	.10	.01	4.6*	1, 423
Ethnicity						
Success with writing process	1 Issue/1 Arg.	1. Essay score (ES)	.04	.04	19.7***	1, 790
		2. ES, gender	.08	.03	5.3***	4, 786
		3. ES, gender, ES x gender	.09	.01	2.5*	4, 782
Success with various kinds of writing	1 Issue/1 Arg.	1. Essay score (ES)	.07	.07	57.4***	1, 790
		2. ES, gender	.11	.04	9.0***	4, 786
		3. ES, gender, ES x gender	.14	.03	6.9***	4, 782
Self-comparison with peers	2 Arg. essays	1. Essay score (ES)	.02	.02	7.0**	1, 398
		2. ES, gender	.02	.01	0.8	4, 394
		3. ES, gender, ES x gender	.05	.03	2.9*	4, 390
Accomplishments in writing	1 Issue/1 Arg.	1. Essay score (ES)	.00	.00	1.9	1, 790
		2. ES, gender	.03	.03	5.4***	4, 786
		3. ES, gender, ES x gender	.04	.02	3.0*	4, 782
Writing sample	2 Issue essays	1. Essay score (ES)	.14	.14	67.2**	1, 425
		2. ES, gender	.18	.04	5.8***	4, 421
		3. ES, gender, ES x gender	.20	.02	2.5*	4, 417
Writing Sample	2 Arg. essays	1. Essay score (ES)	.12	.12	53.2***	1, 398
		2. ES, gender	.13	.01	0.9	4, 394
		3. ES, gender, ES x gender	.16	.04	4.0**	4, 390

Table continued on next page

TABLE 12 (continued)

Group/Indicator	GRE Essay Combination	Independent Variables	Cumulative R^2	Increase in R^2	F	df
English best language						
Accomplishments in writing	2 Issue essays	1. Essay score (ES)	.01	.01	5.2*	1, 425
		2. ES, gender	.01	.00	0.7	1, 424
		3. ES, gender, ES x gender	.03	.01	6.3*	1, 423
Undergraduate major						
Accomplishments in writing	2 Issue essays	1. Essay score (ES)	.01	.01	5.2*	1, 425
		2. ES, gender	.07	.06	4.5***	6, 419
		3. ES, gender, ES x gender	.10	.03	2.4*	6, 413

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

TABLE 13

Increase in R^2 from Interaction Terms in a Hierarchical Regression Analysis

Group/Indicator	1 Issue/ 1 Argument Essay	2 Issue Essays	2 Argument Essays
Gender			
GPA in writing courses	.00	.01*	.00
Success with writing process	.00	.00	.00
Success with various kinds of writing	.00	.01*	.00
Self-comparison with peers	.00	.00	.00
Accomplishments in writing	.00	.01	.00
Writing samples	.00	.00	.01
Ethnicity			
GPA in writing courses	.01	.00	.01
Success with writing process	.01*	.01	.01
Success with various kinds of writing	.03***	.01	.01
Self-comparison with peers	.01	.01	.03*
Accomplishments in writing	.02*	.01	.01
Writing samples	.01	.02*	.04**
English best language			
GPA in writing courses	.00	.00	.00
Success with writing process	.00	.00	.00
Success with various kinds of writing	.00	.00	.00
Self-comparison with peers	.00	.00	.01
Accomplishments in writing	.00	.01*	.00
Writing samples	.00	.00	.00
Undergraduate major			
GPA in writing courses	.01	.02	.01
Success with writing process	.00	.00	.01
Success with various kinds of writing	.00	.01	.01
Self-comparison with peers	.01	.01	.01
Accomplishments in writing	.01	.03*	.01
Writing samples	.00	.02	.02

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

TABLE 14

Performance on GRE Essays (One Issue plus One Argument) versus Performance on Nontest Indicators for Men and Women

		GRE Essay Score Level*					
		1.5*	2.5	3.25	4.0	4.75	5.5
				Means			
Sample X							
	Male	3.2	3.6	3.8	4.0	4.3	4.6
	Female	3.3	3.3	3.6	4.0	4.5	4.7
Sample Y							
	Male	2.7	3.2	3.6	3.9	4.0	4.1
	Female	2.8	3.2	3.4	3.6	3.9	4.2
GPA in writing courses							
	Male	4.5	5.3	5.5	5.9	6.1	6.6
	Female	5.2	5.3	5.5	5.9	6.1	6.5
GPA overall							
	Male	4.7	5.2	5.4	5.5	5.6	6.1
	Female	4.7	5.0	5.0	5.3	5.6	6.1
GPA in major							
	Male	5.1	5.6	5.8	5.8	5.9	6.0
	Female	5.3	5.6	5.4	5.6	5.8	6.3
Self-comparison with peers							
	Male	3.3	3.6	3.9	4.0	4.3	4.3
	Female	3.1	3.6	3.6	3.9	3.9	4.2
Success with writing process							
	Male	3.6	3.8	3.8	3.9	4.1	4.4
	Female	3.8	3.8	3.9	4.0	4.1	4.3
Success with various kinds of writing							
	Male	3.5	3.6	3.7	3.9	4.1	4.2
	Female	3.6	3.7	3.9	3.9	3.9	4.2
Accomplishments in writing							
	Male	2.4	2.2	1.7	1.8	2.2	2.9
	Female	1.7	2.2	1.5	1.8	2.3	2.2
				Percent 5 or 6			
Sample X							
	Male	14	22	31	37	36	67
	Female	5	22	15	27	46	70
Sample Y							
	Male	10	10	20	31	34	36
	Female	0	7	14	20	23	38

Table continued on next page

TABLE 14 (continued)

		GRE Essay Score Level*					
		1.5*	2.5	3.25	4.0	4.75	5.5
		Percent A+					
GPA in writing courses							
	Male	5	17	19	32	41	67
	Female	5	13	22	29	42	67
GPA overall							
	Male	9	15	19	29	20	40
	Female	10	14	13	16	23	41
GPA in major							
	Male	17	25	29	36	37	50
	Female	26	28	20	29	29	50
		Percent Above Average					
Self-comparison with peers							
	Male	35	59	76	80	88	88
	Female	25	57	56	74	77	86

*Score levels represent the midpoint of score ranges.
Note. Figures are based on 354 males and 532 females. Ns by GRE essay score level (1.5 to 5.5, respectively) are as follows:
Male: 20, 55, 118, 87, 58, and 16
Female: 20, 77, 160, 170, 82, and 22

TABLE 15

Performance on GRE Essays (One Issue plus One Argument) versus Performance on Nontest Indicators by Ethnicity

		GRE Essay Score Level*					
		1.5*	2.5	3.25	4.0	4.75	5.5
		Means					
Sample X							
	White	4.2	3.9	4.0	4.0	4.4	4.7
	Black	3.2	3.2	3.4	3.7	4.1	—
	Asian	3.1	2.9	3.5	4.0	4.6	3.5
	Hispanic	3.0	3.0	3.7	4.1	4.0	4.0
	Other	3.0	4.0	3.2	4.3	4.6	5.5
Sample Y							
	White	3.2	3.3	3.7	3.7	4.0	4.3
	Black	3.1	3.2	3.1	3.6	3.4	—
	Asian	2.2	2.9	3.5	3.8	3.9	3.7
	Hispanic	2.0	3.4	3.3	4.1	4.0	3.0
	Other	3.0	3.9	3.2	3.6	4.4	3.5
GPA in writing courses							
	White	4.8	5.0	5.5	5.9	6.1	6.6
	Black	4.7	5.5	5.5	5.7	5.5	—
	Asian	5.1	5.5	5.5	6.0	6.3	6.3
	Hispanic	4.0	5.0	5.3	6.2	5.7	—
	Other	6.0	5.5	5.7	6.2	6.4	6.0
GPA overall							
	White	4.0	4.7	5.2	5.4	5.6	6.1
	Black	4.1	4.8	4.9	5.0	4.8	—
	Asian	5.6	5.9	5.4	5.6	5.9	6.3
	Hispanic	3.5	4.0	5.3	5.4	4.7	—
	Other	6.0	5.2	4.7	5.3	5.6	5.5
GPA in major							
	White	4.4	5.4	5.6	5.7	5.9	6.3
	Black	4.6	5.2	5.2	5.0	5.0	—
	Asian	5.9	6.2	6.0	5.9	5.9	6.7
	Hispanic	5.0	5.0	5.5	6.2	5.7	3.0
	Other	6.0	5.6	5.4	6.1	6.0	5.0
Self-comparison with peers							
	White	3.6	3.5	3.7	3.9	4.1	4.3
	Black	3.3	3.9	3.9	3.7	3.8	—
	Asian	3.2	3.5	3.5	3.9	3.9	4.3
	Hispanic	2.5	3.6	3.4	3.7	3.7	3.0
	Other	3.0	3.7	4.1	4.4	4.4	3.5

Table continued on next page

TABLE 15 (continued)

		GRE Essay Score Level*					
		1.5*	2.5	3.25	4.0	4.75	5.5
				Means			
Success with writing process							
	White	3.8	3.7	3.8	4.0	4.0	4.4
	Black	3.7	4.1	4.1	3.9	3.9	—
	Asian	3.7	3.8	3.7	3.8	4.2	4.2
	Hispanic	3.0	3.4	3.9	4.3	4.2	3.7
	Other	3.7	3.7	3.8	3.9	4.5	3.8
Success with various kinds of writing							
	White	3.8	3.5	3.8	3.9	4.0	4.3
	Black	3.7	4.1	4.0	3.9	3.8	—
	Asian	3.4	3.6	3.5	3.8	4.0	4.2
	Hispanic	2.8	3.2	3.7	4.2	4.0	3.7
	Other	3.9	3.4	3.8	4.1	4.5	4.0
Accomplishments in writing							
	White	1.2	1.6	1.3	1.7	2.3	2.6
	Black	2.0	2.4	1.6	2.0	2.4	—
	Asian	2.1	3.0	2.2	1.8	2.8	2.3
	Hispanic	2.5	1.4	1.8	3.3	1.7	0.0
	Other	4.0	2.0	1.1	1.7	1.0	1.5
				Percent 5 or 6			
Sample X							
	White	60	34	27	32	42	70
	Black	6	21	17	17	33	—
	Asian	0	11	16	32	50	50
	Hispanic	0	0	22	22	0	0
	Other	0	33	9	56	60	100
Sample Y							
	White	0	9	22	24	29	38
	Black	12	3	9	17	0	—
	Asian	0	8	19	23	29	33
	Hispanic	0	20	0	44	33	0
	Other	0	11	11	12	40	50

Table continued on next page

TABLE 15 (continued)

		GRE Essay Score Level*					
		1.5*	2.5	3.25	4.0	4.75	5.5
				Percent A+			
GPA in writing courses							
	White	0	12	20	28	41	68
	Black	6	21	23	24	25	—
	Asian	6	16	21	39	50	67
	Hispanic	0	0	17	44	50	0
	Other	0	12	20	44	60	50
GPA overall							
	White	0	11	17	16	22	41
	Black	6	6	12	12	8	—
	Asian	18	28	23	40	36	33
	Hispanic	0	0	11	22	0	0
	Other	0	22	0	44	40	50
GPA in major							
	White	0	29	22	30	35	50
	Black	12	14	13	12	0	—
	Asian	41	41	38	46	50	67
	Hispanic	0	0	33	44	25	0
	Other	0	25	18	44	20	50
				Percent Above Average			
Self-comparison with peers							
	White	40	45	67	81	85	91
	Black	33	75	71	56	73	—
	Asian	29	51	52	69	71	100
	Hispanic	0	80	47	56	50	0
	Other	0	67	80	100	80	50

*Score levels represent the midpoint of score ranges.

Note. Figures are based on 490 White, 170 Black, 167 Asian, 39 Hispanic, and 37 other participants. Ns by GRE essay score level (1.5 to 5.5, respectively) are as follows:

White: 5, 44, 132, 167, 105, and 32
Black: 15, 36, 69, 34, 11, and 0
Asian: 17, 39, 52, 39, 14, and 3
Hispanic: 2, 4, 18, 9, 3, and 1
Other: 1, 9, 10, 8, 5, and 2

TABLE 16

Performance on GRE Essays (One Issue plus One Argument) versus Performance on Nontest Indicators by Language

		GRE Essay Score Level*					
		1.5*	2.5	3.25	4.0	4.75	5.5
		Means					
Sample X							
	English	3.4	3.5	3.7	4.0	4.4	4.7
	Other	2.9	3.2	3.5	4.0	4.3	4.0
Sample Y							
	English	2.9	3.2	3.5	3.7	3.9	4.2
	Other	2.4	3.2	3.6	3.5	4.0	4.0
GPA in writing courses							
	English	4.9	5.2	5.5	5.9	6.1	6.5
	Other	4.9	5.6	5.5	5.9	6.1	7.0
GPA overall							
	English	4.3	4.9	5.1	5.4	5.5	6.1
	Other	5.5	5.8	5.5	5.4	6.0	7.0
GPA in major							
	English	5.0	5.4	5.5	5.6	5.8	6.1
	Other	5.6	6.2	5.8	6.0	5.9	7.0
Self-comparison with peers							
	English	3.3	3.7	3.7	3.9	4.1	4.2
	Other	3.1	3.3	3.5	3.6	4.1	—
Success with writing process							
	English	3.8	3.8	3.9	4.0	4.1	4.3
	Other	3.5	3.8	3.7	3.8	4.1	4.7
Success with various kinds of writing							
	English	3.7	3.7	3.8	3.9	4.0	4.2
	Other	3.3	3.7	3.6	3.7	4.0	4.1
Accomplishments in writing							
	English	1.6	2.1	1.6	1.8	2.3	2.5
	Other	2.8	2.5	1.6	1.9	2.3	4.0

Table continued on next page

TABLE 16 (continued)

		GRE Essay Score Level*					
		1.5*	2.5	3.25	4.0	4.75	5.5
		Percent 5 or 6					
Sample X							
	English	15	23	23	31	43	71
	Other	0	20	16	27	35	0
Sample Y							
	English	8	7	15	23	26	38
	Other	0	10	22	24	35	0
		Percent A+					
GPA in writing courses							
	English	7	13	22	30	40	66
	Other	0	22	12	32	53	100
GPA overall							
	English	3	11	15	20	19	39
	Other	21	27	23	20	42	100
GPA in major							
	English	14	20	22	29	30	49
	Other	36	50	31	46	47	100
		Percent Above Average					
Self-comparison with peers							
	English	30	61	66	76	82	87
	Other	31	47	55	72	79	0

*Score levels represent the midpoint of score ranges.

Note. Figures are based on 762 participants whose best language was English and 142 whose best language was other than English. Ns by GRE essay score level (1.5 to 5.5, respectively) are as follows:

English: 27, 101, 237, 227, 121, and 37

Other: 13, 32, 44, 30, 19, and 1

TABLE 17

Performance on GRE Essays (One Issue plus One Argument) versus Performance on Nontest Indicators by Major

		GRE Essay Score Level*					
		1.5*	2.5	3.25	4.0	4.75	5.5
		Means					
Sample X							
	Natural Science	3.6	3.7	3.7	4.1	4.7	4.4
	Engineering	2.7	3.6	3.8	3.9	4.1	5.5
	Social Science	3.4	3.5	3.6	4.0	4.3	4.7
	Humanities	3.7	3.8	4.0	4.2	4.6	4.9
	Education	—	3.0	3.6	4.0	4.3	4.0
	Business	2.5	3.4	3.5	3.7	4.2	5.5
	Other	3.5	2.8	3.7	3.7	4.1	4.3
Sample Y							
	Natural Science	2.6	3.4	3.5	3.7	4.0	3.9
	Engineering	2.3	2.8	3.4	3.7	3.7	4.0
	Social Science	2.8	3.4	3.5	3.7	4.4	3.8
	Humanities	4.0	4.0	3.6	3.7	4.0	4.9
	Education	—	3.5	3.6	3.9	3.9	4.0
	Business	2.7	2.9	3.0	3.6	3.7	4.7
	Other	3.0	2.8	3.7	3.6	3.7	4.3
GPA in writing courses							
	Natural Science	5.3	5.2	5.5	6.0	5.7	6.4
	Engineering	4.5	5.9	5.6	6.1	6.3	7.0
	Social Science	4.9	5.2	5.5	5.7	6.3	6.3
	Humanities	3.7	5.6	5.4	6.0	6.3	6.8
	Education	—	5.3	5.6	5.9	6.1	7.0
	Business	5.0	5.0	5.7	5.8	5.6	7.0
	Other	4.2	5.5	5.6	5.9	5.9	6.7
GPA overall							
	Natural Science	5.0	5.0	5.2	5.5	5.4	6.0
	Engineering	5.6	6.0	5.2	5.6	5.9	6.5
	Social Science	4.4	5.0	5.1	5.2	5.7	6.0
	Humanities	3.0	5.0	5.0	5.8	5.6	6.1
	Education	—	5.0	5.1	5.1	5.6	5.0
	Business	4.5	4.8	5.2	4.8	5.6	7.0
	Other	3.7	4.8	5.2	5.3	5.3	6.0

Table continued on next page

TABLE 17 (continued)

		GRE Essay Score Level*					
		1.5*	2.5	3.25	4.0	4.75	5.5
		Means					
GPA in major							
	Natural Science	5.1	5.5	5.4	5.5	5.5	5.9
	Engineering	5.6	6.2	5.5	5.6	5.8	6.5
	Social Science	4.9	5.7	5.5	5.7	6.0	6.0
	Humanities	3.7	5.6	5.8	6.3	6.2	6.7
	Education	—	5.7	5.8	5.9	6.0	5.0
	Business	5.2	5.4	5.4	5.2	6.0	6.7
	Other	5.7	5.3	5.7	5.9	5.4	6.0
Self-comparison with peers							
	Natural Science	3.1	3.6	3.7	3.9	4.3	4.4
	Engineering	3.1	3.6	3.7	4.0	4.4	4.5
	Social Science	3.7	3.6	3.6	3.8	4.1	3.9
	Humanities	3.3	3.8	3.8	3.9	4.2	4.0
	Education	—	3.4	3.7	3.9	4.0	4.0
	Business	3.0	3.6	3.9	3.8	3.9	4.7
	Other	3.2	3.8	3.6	4.1	3.6	4.7
Success with writing process							
	Natural Science	3.8	3.9	3.8	3.9	4.1	4.3
	Engineering	3.5	3.9	3.8	4.0	4.3	4.3
	Social Science	4.0	3.8	3.9	3.9	4.2	4.2
	Humanities	3.5	3.9	4.0	4.0	4.1	4.3
	Education	—	2.8	3.8	3.9	3.9	4.7
	Business	3.3	3.6	3.9	3.9	3.7	4.8
	Other	3.5	4.0	3.8	4.1	3.9	4.5
Success with various kinds of writing							
	Natural Science	3.6	3.8	3.7	3.9	3.9	4.2
	Engineering	3.6	3.8	3.7	3.9	4.1	4.7
	Social Science	4.0	3.6	3.8	3.9	4.1	4.2
	Humanities	3.2	3.6	4.0	3.9	4.0	4.1
	Education	—	3.6	4.0	4.0	4.0	4.5
	Business	3.3	3.3	3.8	3.7	3.9	4.6
	Other	3.5	3.8	3.7	4.0	3.8	4.4

Table continued on next page

TABLE 17 (continued)

		GRE Essay Score Level*					
		1.5*	2.5	3.25	4.0	4.75	5.5
				Means			
Accomplishments in writing							
	Natural Science	2.6	2.1	1.3	1.8	2.3	2.4
	Engineering	0.3	2.8	2.0	2.1	2.9	2.0
	Social Science	2.0	1.6	1.3	1.5	2.1	2.7
	Humanities	2.0	2.2	2.0	2.4	2.5	1.4
	Education	—	1.5	1.4	1.3	2.1	0.0
	Business	2.0	2.2	1.4	2.1	1.2	4.0
	Other	3.0	3.2	2.2	2.1	1.9	4.3
				Percent 5 or 6			
Sample X							
	Natural Science	0	34	22	35	58	70
	Engineering	14	20	29	24	33	100
	Social Science	13	28	20	33	46	60
	Humanities	33	40	32	36	46	71
	Education	—	0	7	27	50	0
	Business	0	15	23	21	20	100
	Other	25	14	13	22	20	67
Sample Y							
	Natural Science	0	18	10	22	31	33
	Engineering	0	5	20	26	18	0
	Social Science	14	0	30	24	42	20
	Humanities	50	20	28	27	35	57
	Education	—	0	21	20	12	0
	Business	0	8	0	22	10	67
	Other	0	0	21	22	19	67
				Percent A+			
GPA in writing courses							
	Natural Science	13	17	21	32	35	55
	Engineering	0	32	23	40	55	100
	Social Science	0	17	17	21	46	50
	Humanities	0	20	14	32	50	83
	Education	—	14	31	29	50	100
	Business	0	0	24	32	20	100
	Other	0	7	24	28	28	67

Table continued on next page

Table 17 (continued)

		GRE Essay Score Level*					
		1.5*	2.5	3.25	4.0	4.75	5.5
				Percent A+			
GPA overall							
	Natural Science	20	16	18	19	19	45
	Engineering	12	35	19	40	41	50
	Social Science	0	10	12	19	19	30
	Humanities	0	20	10	24	20	29
	Education	—	12	21	13	25	0
	Business	0	8	9	12	10	100
	Other	0	6	16	14	17	33
GPA in major							
	Natural Science	27	36	21	27	23	46
	Engineering	25	52	22	43	36	50
	Social Science	12	30	25	25	35	45
	Humanities	0	40	26	48	40	71
	Education	—	14	23	29	37	0
	Business	25	14	9	24	30	67
	Other	0	0	37	34	28	33
				Percent Above Average			
Self-comparison with peers							
	Natural Science	13	56	63	72	88	91
	Engineering	29	65	68	83	91	100
	Social Science	67	55	59	63	92	70
	Humanities	33	60	68	79	76	86
	Education	—	37	64	73	87	100
	Business	25	57	64	72	80	100
	Other	25	75	61	90	56	100

*Score levels represent the midpoint of score ranges.

Note. Figures are based on 247 natural science majors, 122 engineering majors, 174 social science majors, 99 humanities majors, 46 education majors, 78 business majors, and 102 other majors. Ns by GRE essay score level (1.5 to 5.5, respectively) are as follows:

Natural Sciences: 15, 36, 82, 74, 26, and 11
Engineering: 6, 21, 35, 27, 22, and 2
Social Sciences: 6, 20, 55, 52, 26, and 11
Humanities: 3, 5, 28, 24, 28, and 7
Education: 0, 7, 13, 14, 8, and 1
Business: 4, 14, 22, 25, 10, and 3
Other: 4, 16, 29, 27, 16, and 3

TABLE 18

Comparison of Actual Standing on Three Nontest Indicators of Writing Skill with Standing Predicted from GRE Essay Scores (by Gender)

Indicator	Group	Mean Actual	Mean Predicted	Mean Difference
Writing Sample X	Men	3.91	3.85	+.06
	Women	3.84	3.89	-.04
Writing Sample Y	Men	3.66	3.57	+.09
	Women	3.54	3.60	-.06
GPA in Courses Requiring Writing	Men	5.67	5.69	-.02
	Women	5.73	5.72	+.01

Note. Based on 541 women and 360 men.

TABLE 19

Comparison of Actual Standing on Three Nontest Indicators of Writing Skill with Standing Predicted from GRE Essay Scores (by Ethnicity)

Indicator	Group	Mean Actual	Mean Predicted	Mean Difference
Writing Sample X	White	4.13	4.00	+.14
	Black	3.43	3.68	-.24
	Asian	3.52	3.66	-.15
	Hispanic	3.70	3.87	-.17
Writing Sample Y	White	3.78	3.69	+.09
	Black	3.24	3.43	-.19
	Asian	3.29	3.42	-.14
	Hispanic	3.58	3.59	-.01
GPA in Courses Requiring Writing	White	5.81	5.83	-.03
	Black	5.45	5.53	-.08
	Asian	5.66	5.50	+.16
	Hispanic	5.55	5.72	-.18

Note. Based on 490 White, 170 Black, 143 Asian, and 59 Hispanic test takers.

TABLE 20

Comparison of Actual Standing on Three Nontest Indicators of Writing Skill with Standing Predicted from GRE Essay Scores (by Best Language)

Indicator	Group	Mean Actual	Mean Predicted	Mean Difference
Writing Sample X	English best	3.92	3.90	+.03
	Other best	3.59	3.72	-.13
Writing Sample Y	English best	3.61	3.61	+.00
	Other best	3.44	3.47	-.03
GPA in Courses Requiring Writing	English best	5.72	5.74	-.02
	Other best	5.64	5.56	+.08

Note. Based on 762 examinees whose best language was English and 142 whose best language was not English.

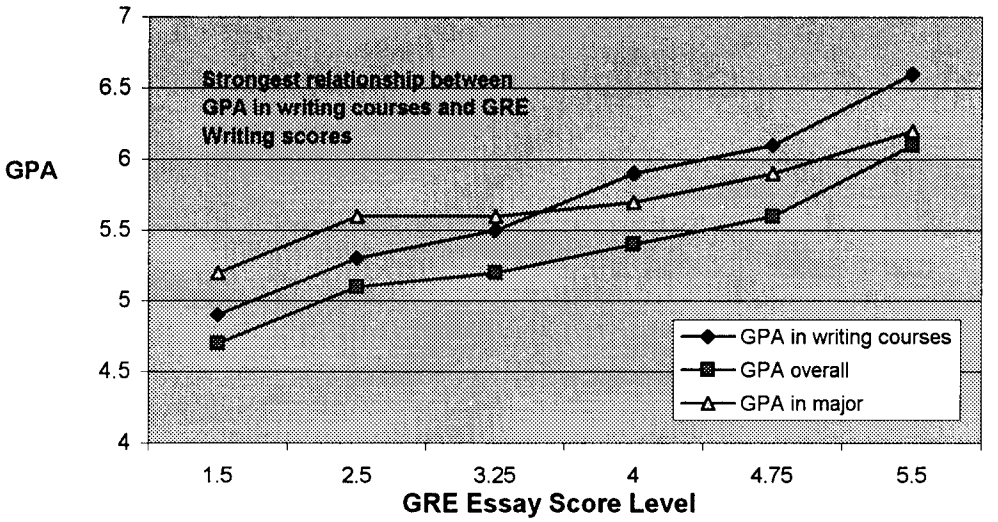


FIGURE 1: Relationship Between GRE Writing Assessment Scores and Undergraduate GPAs

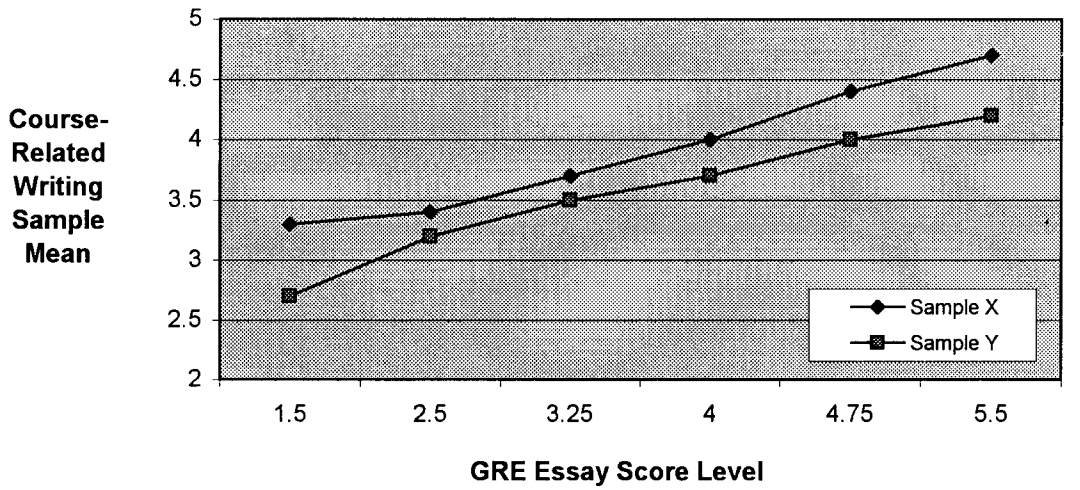


FIGURE 2: Relationship Between GRE Writing Assessment Scores and Course-Related Writing Samples

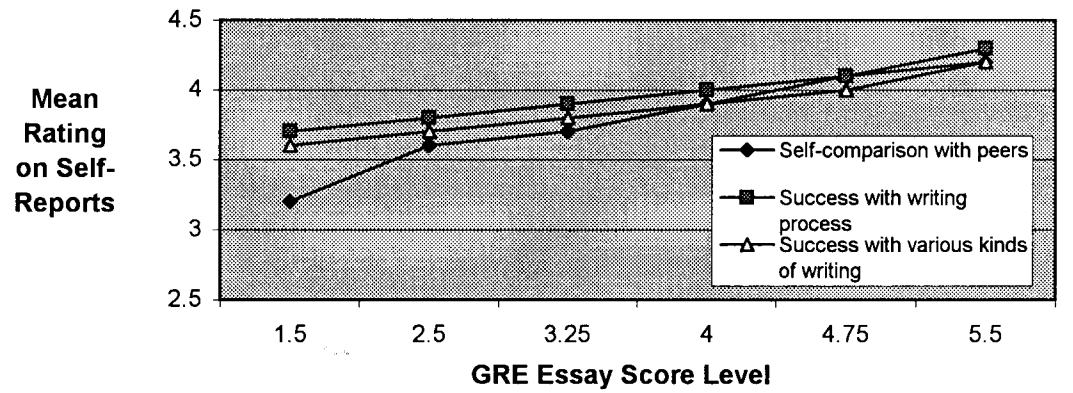


FIGURE 3: Self-Report Ratings versus GRE Writing Assessment Scores

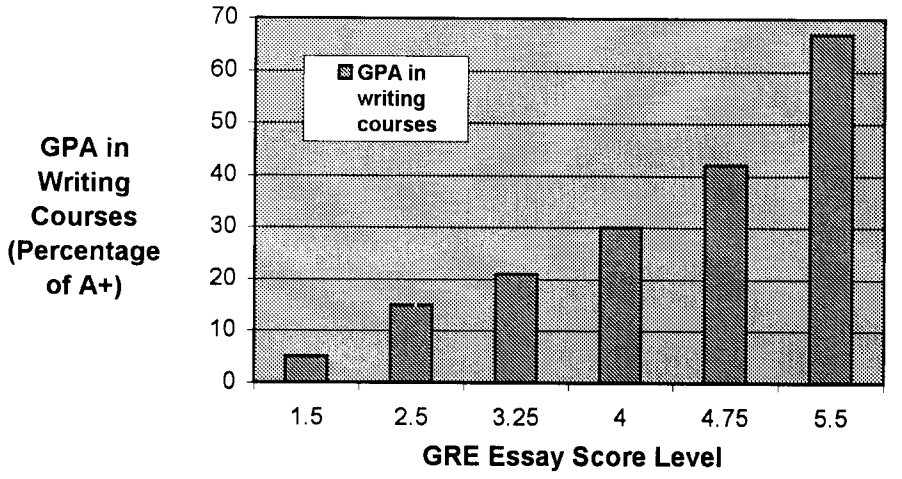


FIGURE 4: GPA in Writing Courses by Performance on GRE Essays

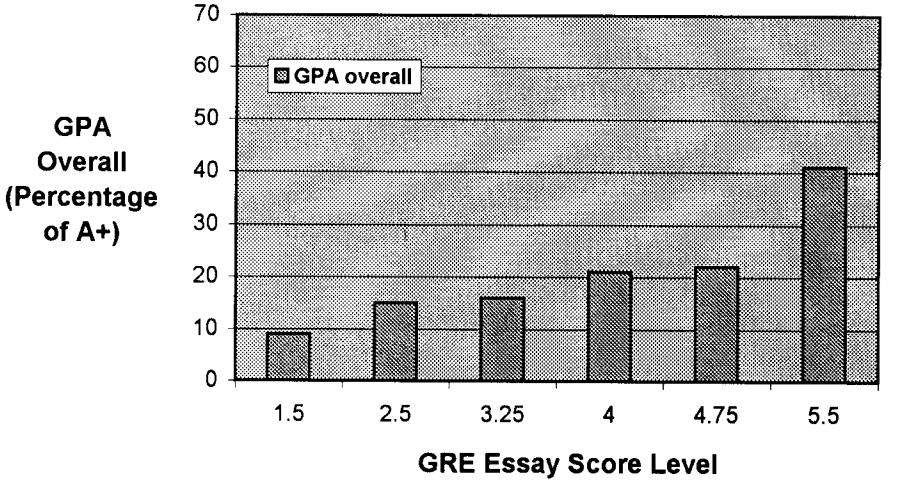


FIGURE 5: GPA Overall by Performance on GRE Essays

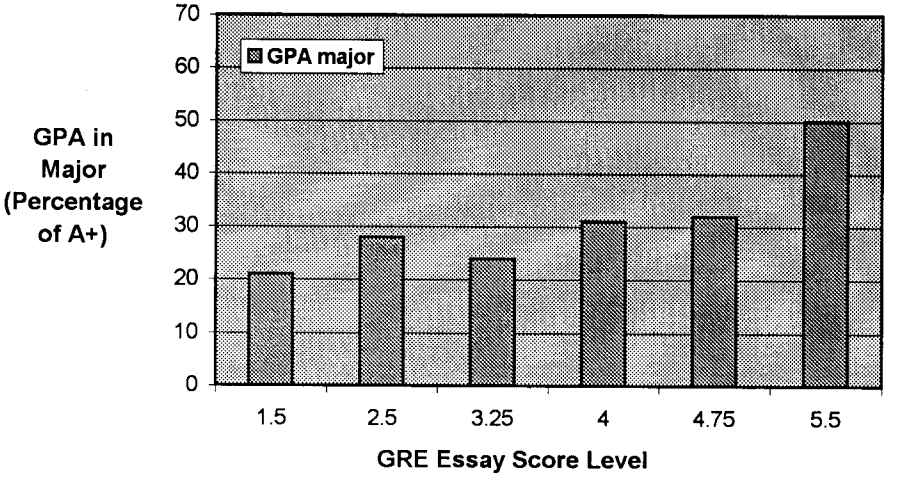


FIGURE 6: GPA in Major by Performance on GRE Essays

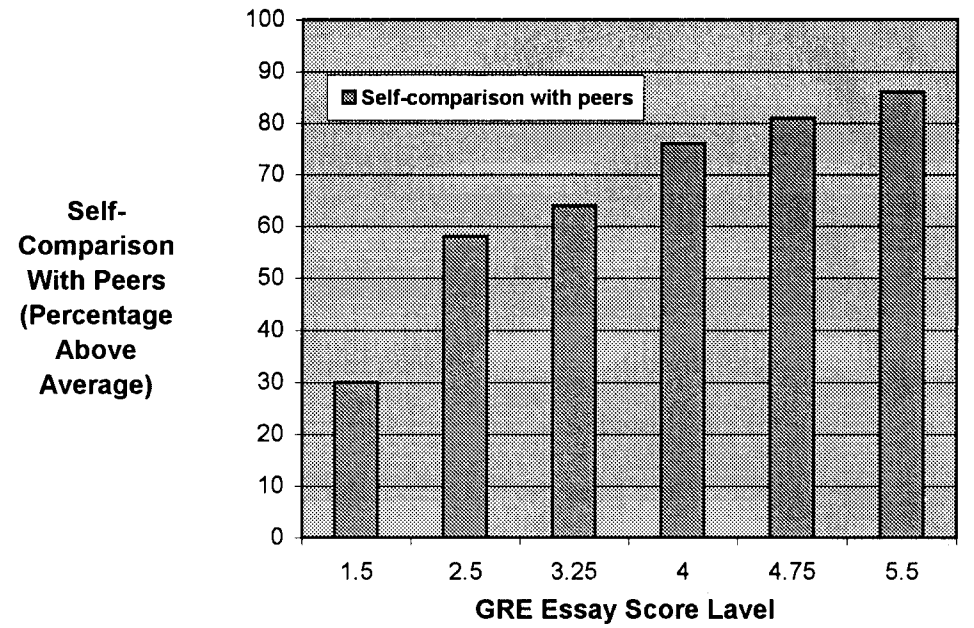


FIGURE 7: Self-Comparison With Peers by Performance on GRE Essays

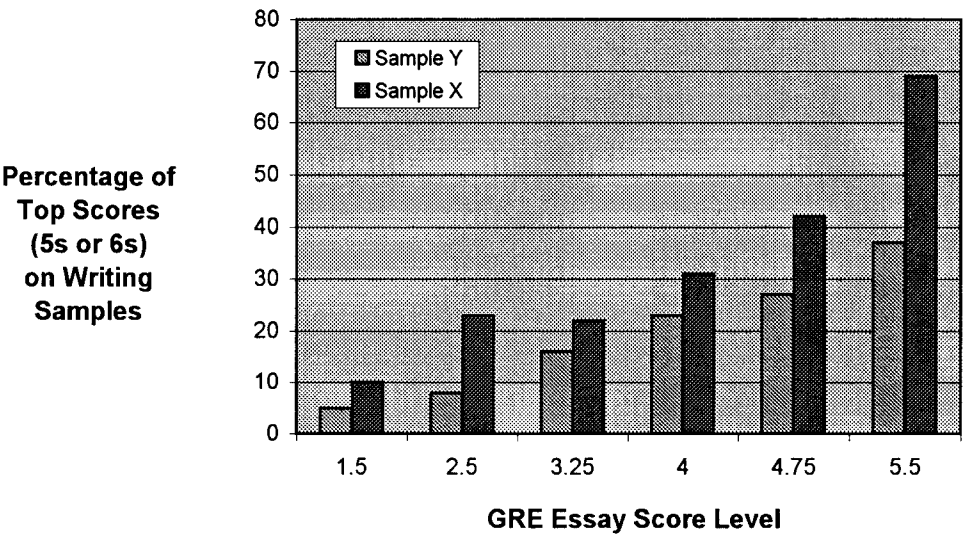
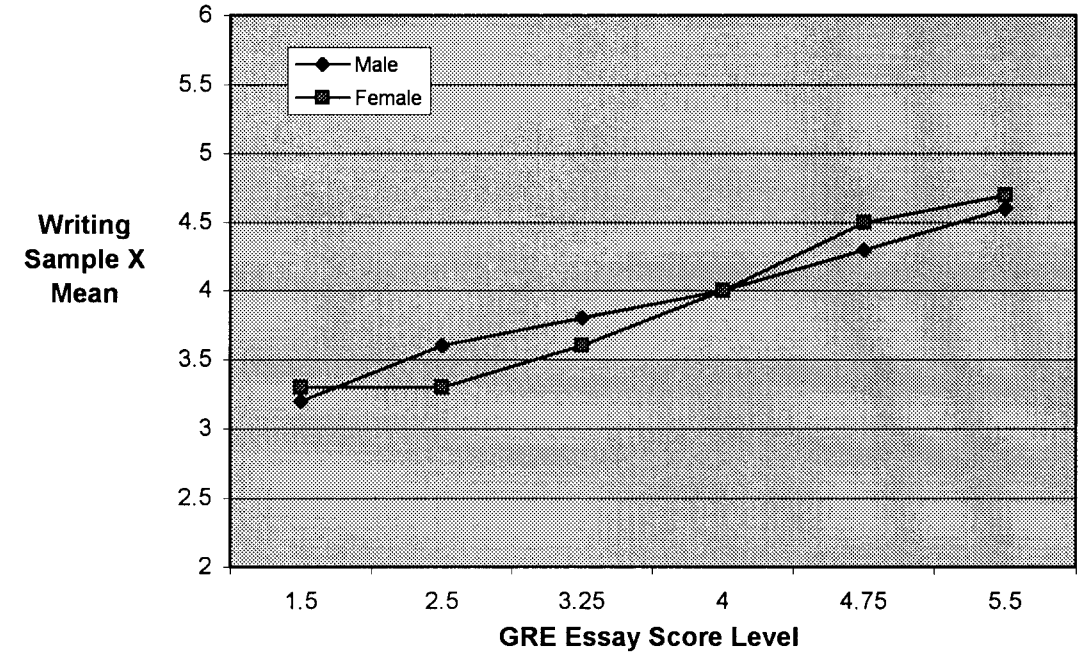


FIGURE 8: Performance on Course-Related Writing Samples by GRE Writing Assessment Level



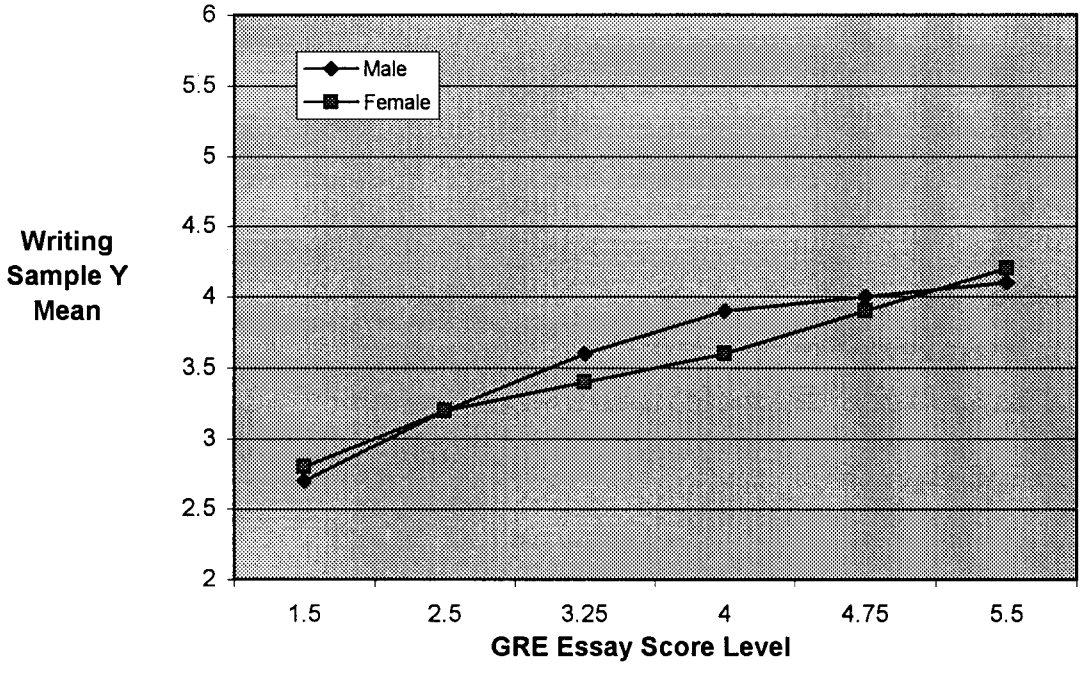
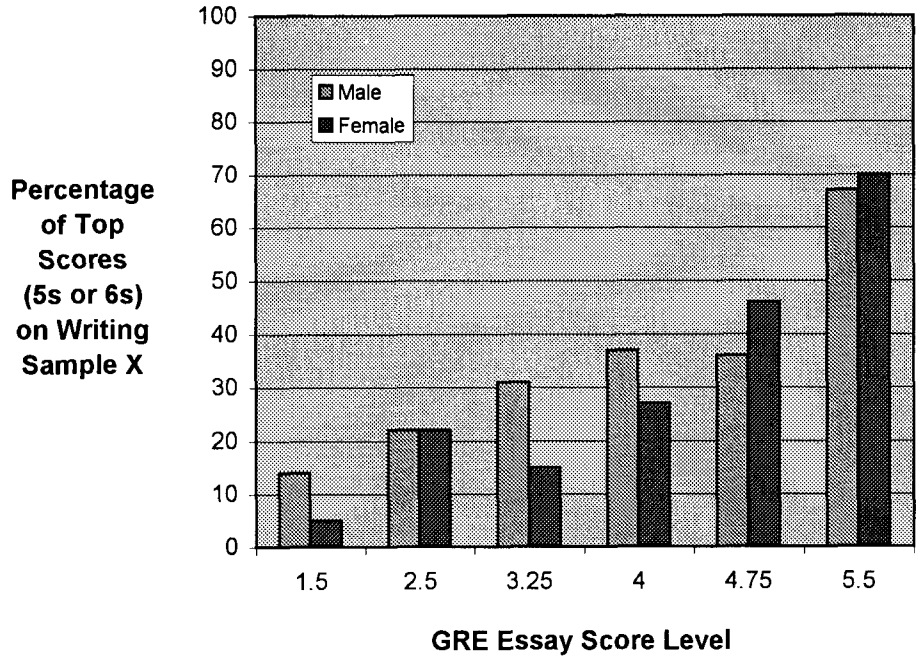


FIGURE 9: Performance on Course-Related Writing Sample X and Y by GRE Writing Assessment Level (by Gender)



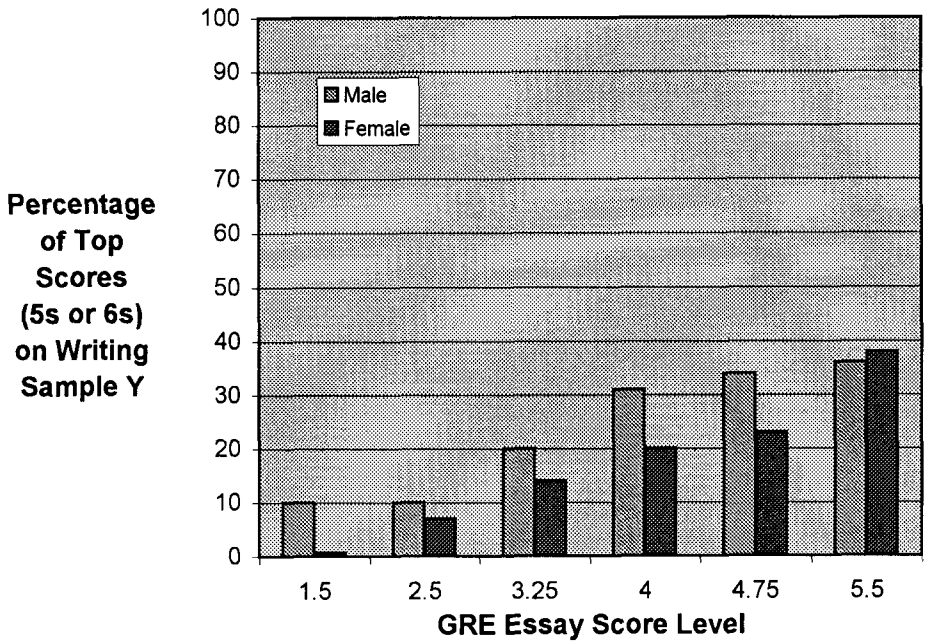
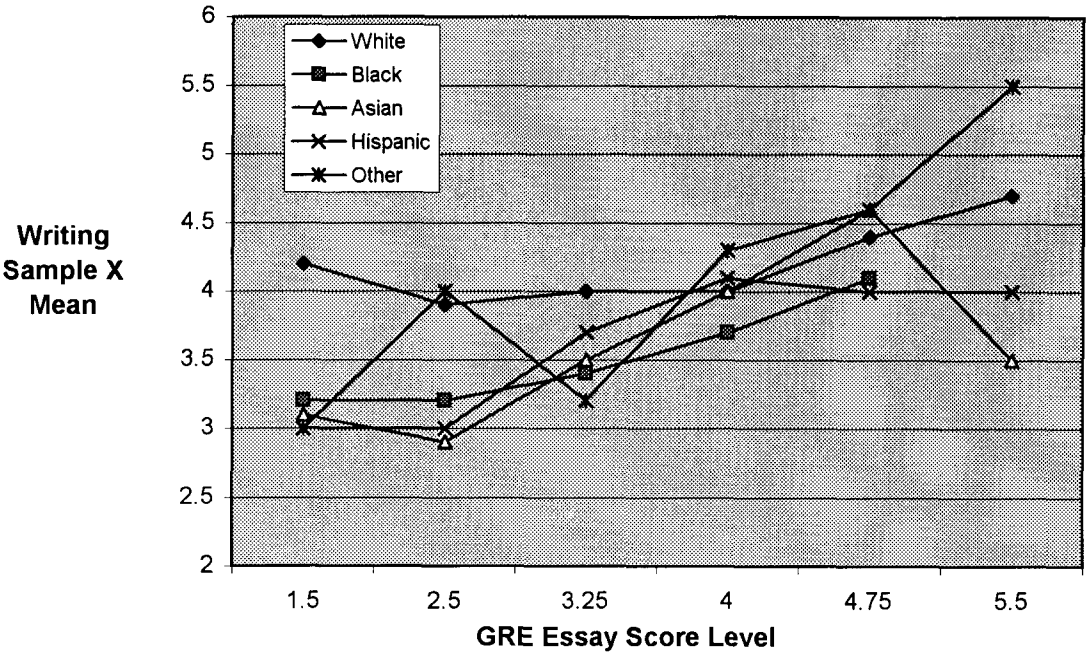


FIGURE 10: Performance on Course-Related Writing Sample X and Y by GRE Writing Assessment Level (by Gender)



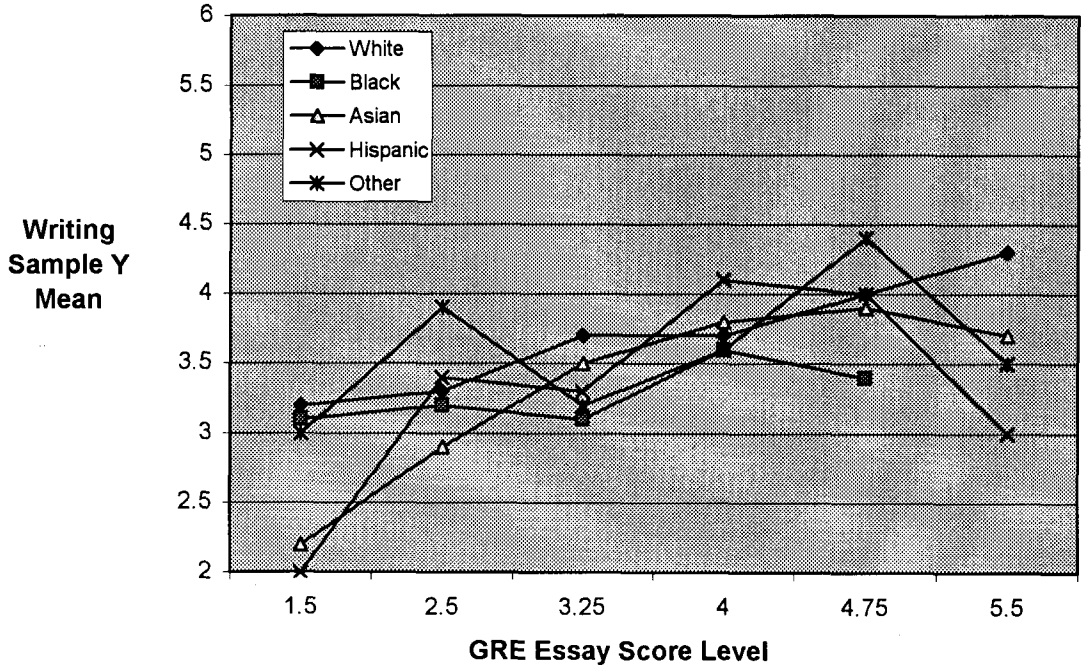
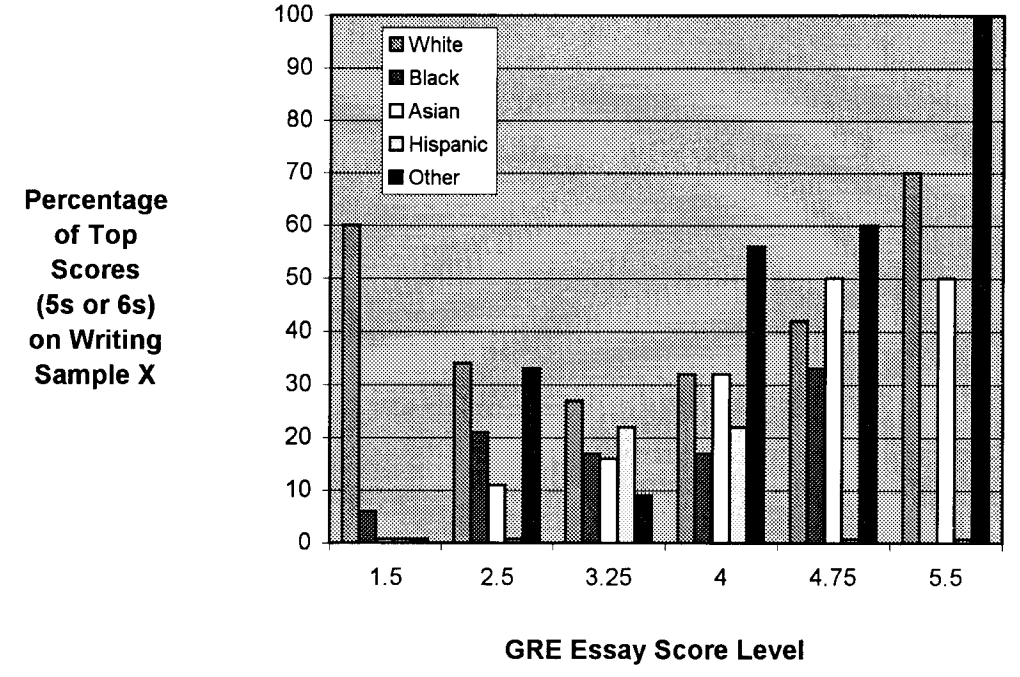


FIGURE 11: Performance on Course-Related Writing Sample X and Y by GRE Writing Assessment Level (by Ethnicity)



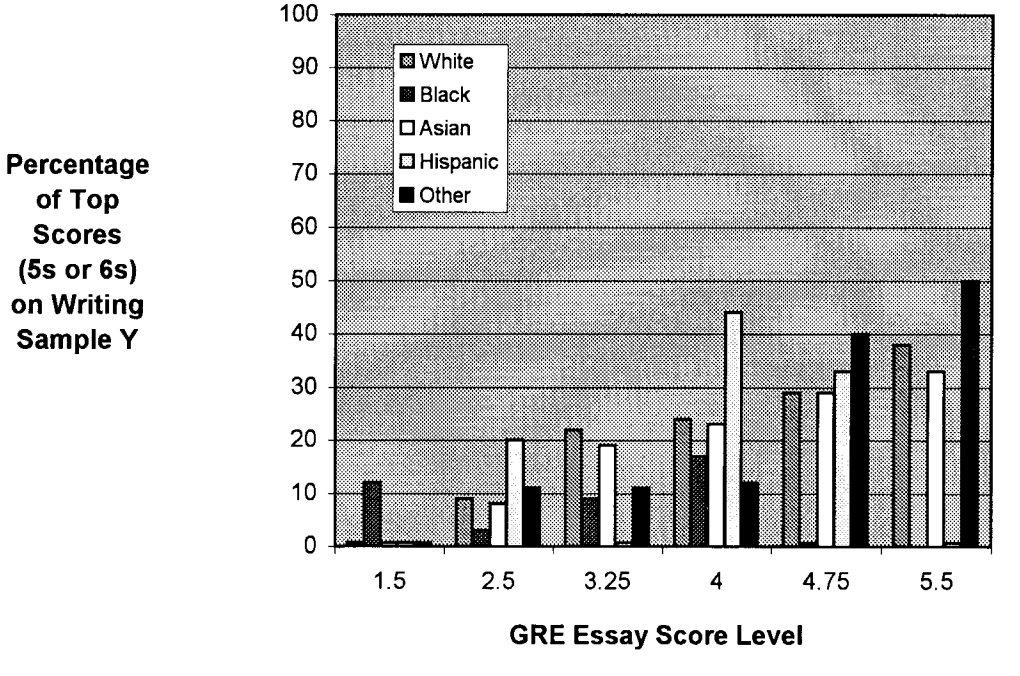
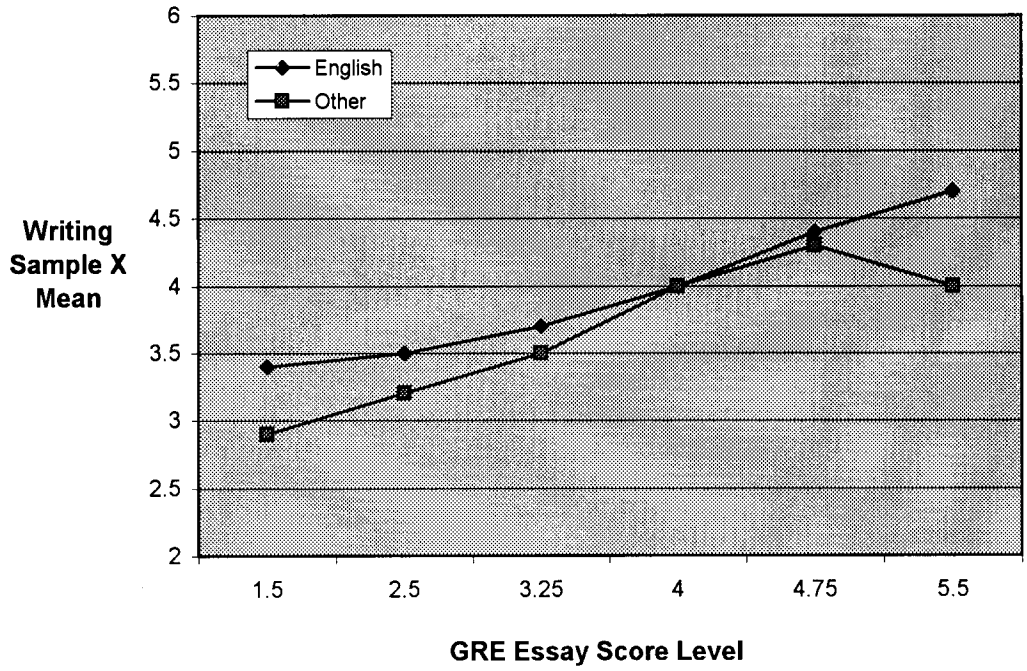


FIGURE 12: Performance on Course-Related Writing Sample X and Y by GRE Writing Assessment Level (by Ethnicity)



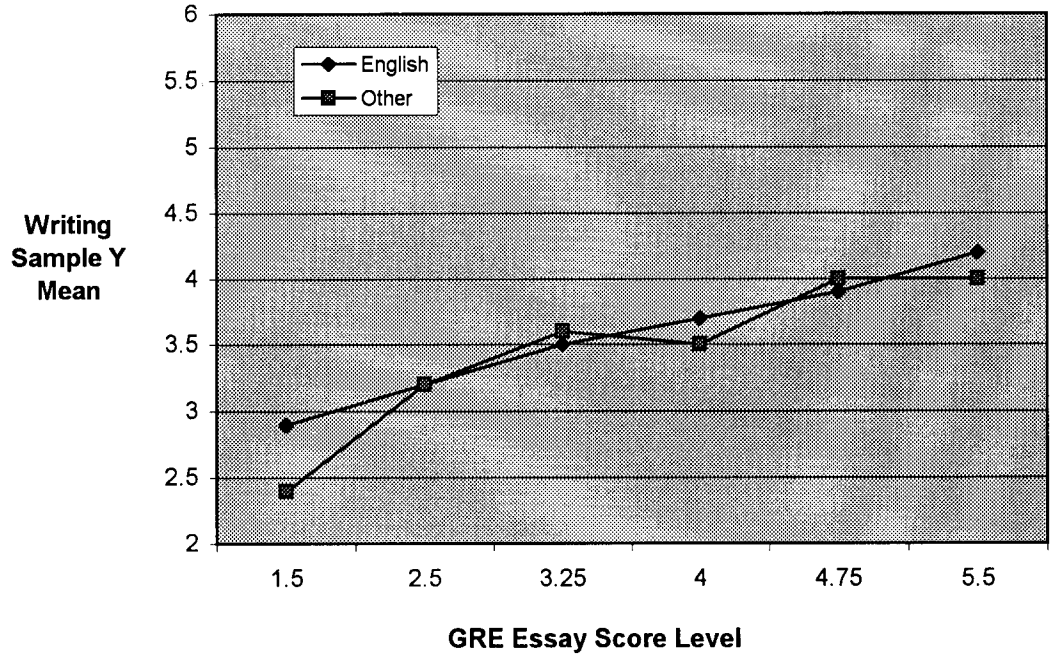
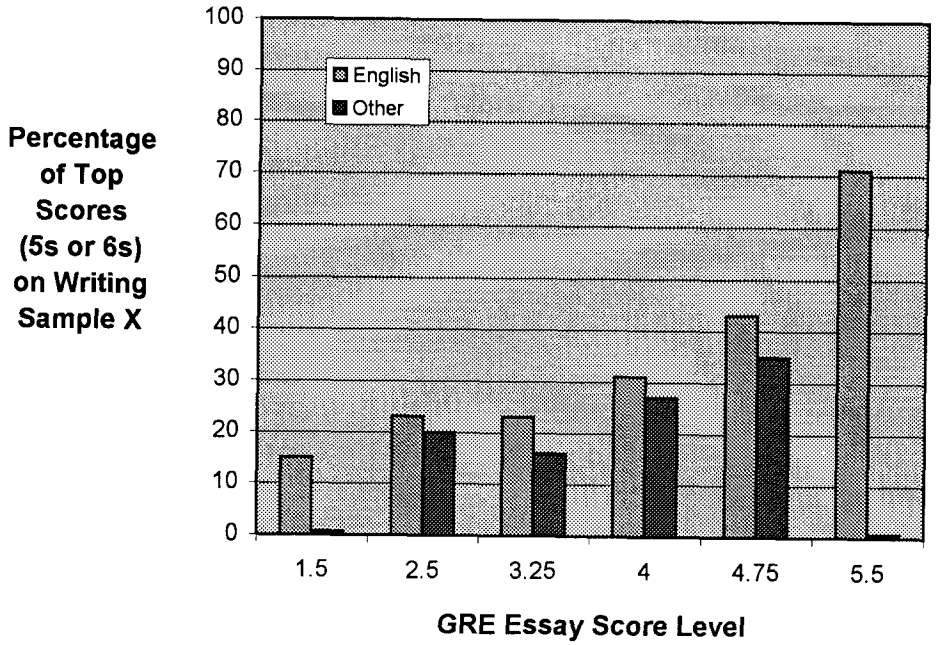


FIGURE 13: Performance on Course-Related Writing Sample X and Y by GRE Writing Assessment Level (by Best Language)



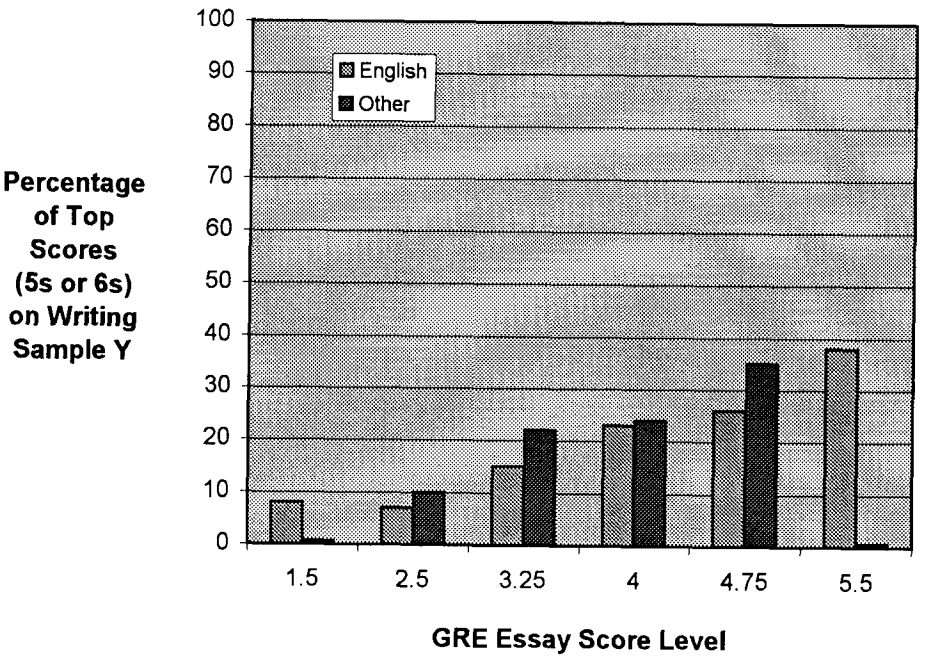
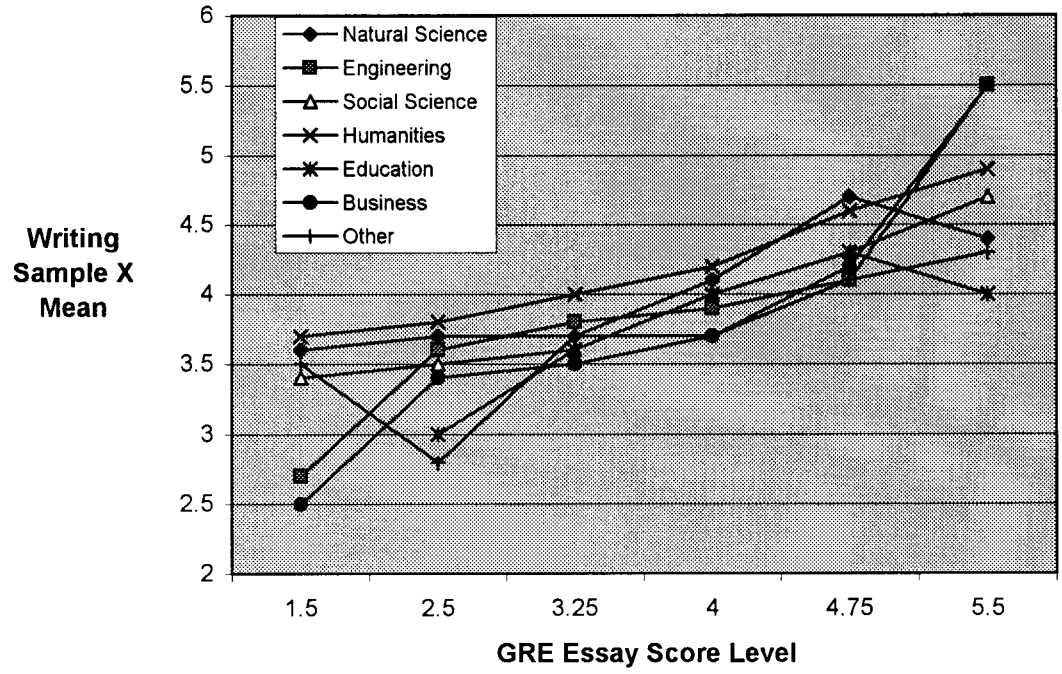


FIGURE 14: Performance on Course-Related Writing Sample X and Y by GRE Writing Assessment Level (by Best Language)



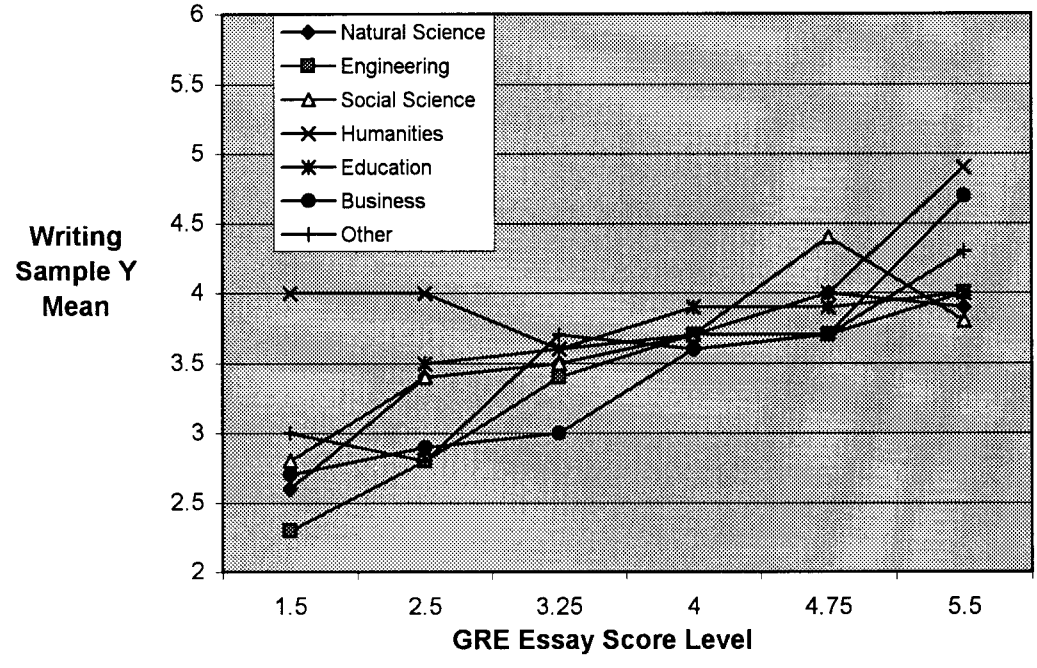
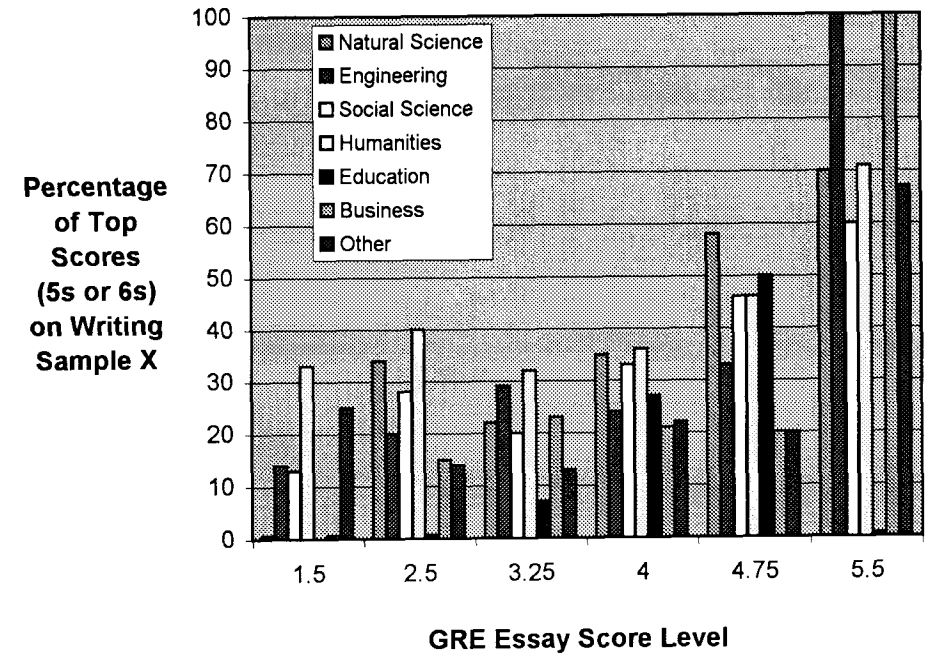


FIGURE 15: Performance on Course-Related Writing Sample X and Y by GRE Writing Assessment Level (by Undergraduate Major)



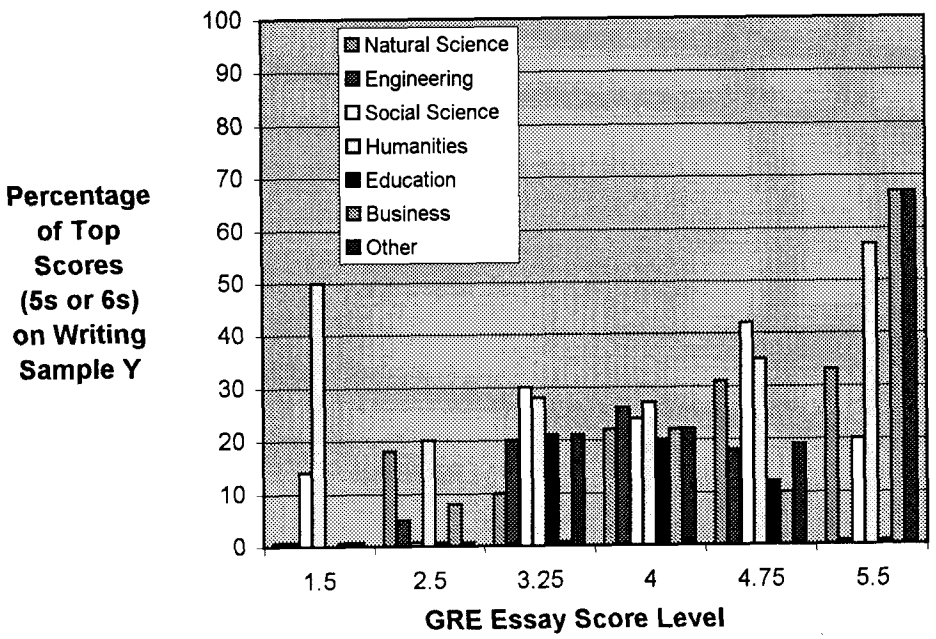


FIGURE 16: Performance on Course-Related Writing Sample X and Y by GRE Writing Assessment Level (by Undergraduate Major)

Appendix A

GRE Writing Study Questionnaire

Your name: _____

GRE WRITING STUDY QUESTIONNAIRE

Before answering the questionnaire, please label your writing samples:

(1) Choose the writing sample that you think is <u>better</u>, and label it "Sample X." (You may interpret "better" according to your own assessment or any evaluation/feedback from other students or instructors.)

(2) On the <u>other</u> sample, write "Sample Y."

(3) Write your name and Social Security number on both Samples X and Y.

About your two course-related writing samples

1. For each sample, provide the name of each course and briefly describe the assignment (as well as the purpose and audience for each sample).

   Sample X: _____

   _____

   Sample Y: _____

   _____

2. Were both samples assigned by the same instructor? No.....1 Yes.....2

For each question below, circle the <u>one</u> number that most closely describes each sample.
(You should end up with a total of 2 numbers only, one for each sample.)

3. Which of the following <u>best</u> describes the course-related writing samples you have provided?

		Sample X	Sample Y
a.	Essay (the statement and development of a proposition or point of view)	1	1
b.	Research paper (based on reference materials).....	2	2
c.	Report of an experiment	3	3
d.	Case study (analysis and presentation of a case situation).....	4	4
e.	Summary or literature review.....	5	5
f.	Review or analysis of a book, article, etc.	6	6
g.	Plan/proposal (describing a way to address an issue or solve a problem)	7	7
h.	Other.....	8	8

Please describe: _____

	Sample X	Sample Y
4. Were the samples written in or outside of class?		
a. Entirely or mostly in class	1	1
b. Entirely or mostly outside of class	2	2
5. How long ago did you write the samples?	Sample X	Sample Y
a. Within the past month	1	1
b. One to twelve months ago	2	2
c. More than one year ago	3	3
6. About how much time from start to finish did you spend composing (not researching) the samples?	Sample X	Sample Y
a. Two hours or less	1	1
b. About 3 to 9 hours	2	2
c. About 10 or more hours	3	3
7. How much information did you retrieve from such sources as the library, Internet, etc.?	Sample X	Sample Y
a. Little if any	1	1
b. Some	2	2
c. A moderate amount	3	3
d. A substantial amount	4	4
8. How much information did you recall from personal observations or experiences?	Sample X	Sample Y
a. Little if any	1	1
b. Some	2	2
c. A moderate amount	3	3
d. A substantial amount	4	4
9. How much did the samples require you to consider various perspectives or contrasting points of view?	Sample X	Sample Y
a. Little if any	1	1
b. Some	2	2
c. A moderate amount	3	3
d. A substantial amount	4	4

10. How much did the samples require you to construct an argument or provide evidence to support your own views or position on an issue?

	Sample X	Sample Y
a. Little if any	1	1
b. Some	2	2
c. A moderate amount	3	3
d. A substantial amount	4	4

11. How much did the samples require you to critique or analyze someone else's argument or position and to evaluate its strengths and weaknesses?

	Sample X	Sample Y
a. Little if any	1	1
b. Some	2	2
c. A moderate amount	3	3
d. A substantial amount	4	4

12. What grade did you receive on the samples?

	A+	A(or A-)	B+	B(or B-)	C+	C(or less)	Ungraded
Sample X.....	6	5	4	3	2	1	0
Sample Y.....	6	5	4	3	2	1	0

13. What grade do you think you should have received on them?

	A+	A(or A-)	B+	B(or B-)	C+	C(or less)
Sample X.....	6	5	4	3	2	1
Sample Y.....	6	5	4	3	2	1

14. How much did these samples determine your course grade?

	Sample X	Sample Y
a. Little if at all	1	1
b. Some	2	2
c. A moderate amount	3	3
d. A substantial amount	4	4

15. How much other feedback (e.g., conferences with the instructor, reviews by other students, comments in margins of your papers) did you receive?

	Sample X	Sample Y
a. Little if any	1	1
b. Some	2	2
c. A moderate amount	3	3
d. A substantial amount	4	4

16. In your opinion, how much did the instructor consider the content or substance of your sample (i.e., your ideas and information) in determining the grade you received?

	Sample X	Sample Y
a. Little if at all	1	1
b. Some	2	2
c. A moderate amount	3	3
d. A substantial amount	4	4

17. In your opinion, how much did the instructor consider such characteristics of your writing as organization, sentence structure, word choice, and punctuation in evaluating your samples?

	Sample X	Sample Y
a. Little if at all	1	1
b. Some	2	2
c. A moderate amount	3	3
d. A substantial amount	4	4

18. Who chose the specific topic or subject on which you wrote?

	Sample X	Sample Y
a. I did	1	1
b. The instructor did	2	2
c. Determined jointly by both the instructor and me	3	3

19. How much help did you get in drafting, revising, and editing the writing samples?

	Sample X	Sample Y
a. Drafting:		
Little or no help	1	1
Moderate or substantial help	2	2
b. Editing/Revising:		
Little or no help	1	1
Moderate or substantial help	2	2

About your writing for college courses in general

20. For each of the activities listed, please indicate how successful you have been in your college courses.

	Extremely successful	Quite successful	Somewhat successful	Not very successful	Not at all successful
a. Thinking about an assignment (e.g., developing ideas, gathering information, deciding on a focus).	5	4	3	2	1
b. Organizing (e.g., making outlines, deciding on the order of ideas).	5	4	3	2	1
c. Revising (e.g., improving sentence phrasing, rearranging ideas, correcting grammar and punctuation, developing or supporting ideas more fully).	5	4	3	2	1

21. For each of the kinds of writing listed, please indicate how successful you have been in your college courses.

	Extremely successful	Quite successful	Somewhat successful	Not very successful	Not at all successful	I did not do this
a. Personal writing (e.g., recording experiences, thoughts, or feelings in a journal or field notebook)	5	4	3	2	1	0
b. Creative writing (e.g., poems or short stories)	5	4	3	2	1	0
c. Persuasive writing (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, work of art, article, theory)	5	4	3	2	1	0
e. Description/explanation (e.g., describing an experiment or a phenomenon)	5	4	3	2	1	0
f. Examinations (e.g., long essay answers)	5	4	3	2	1	0
g. Research papers	5	4	3	2	1	0
h. Applications, resumes, etc.	5	4	3	2	1	0

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

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

About your writing activities and accomplishments

23. Please indicate whether or not you have engaged in the following activities. (*Circle one number for each activity.*)

	Yes	No
a. Wrote a "letter to the editor" that was published.	1	2
b. Wrote a feature article, column, or editorial that was published	1	2
c. Wrote poetry, fiction, or essays that were published.	1	2
d. Was on the editorial staff of a publication or a radio or television station.	1	2
e. Wrote a speech that was given at a large public gathering	1	2
f. Wrote advertising or public relations material for a company/organization.	1	2
g. Wrote technical manuals or other instructional material.	1	2
h. Was invited to participate in a national or regional writer's workshop.	1	2
i. Authored or co-authored a paper for a professional meeting.	1	2
j. Authored or co-authored an article published in a scholarly journal.	1	2

Future interest

24. Would you be interested in participating (for pay), perhaps 6 months to a year from now, in follow-up research on the GRE writing test?

    No.....1 Yes.....2

    If yes, please give us your current address and telephone number.

    Address _____

    Home phone number (_____)_____

    Please give us the address and phone number of a parent, relative, etc. who would best know how to contact you in the future?

    Name _____

    Address _____

    Phone number (_____)_____

25. May we use, as examples in our research reports, the samples of writing you provided? Your identification will be kept confidential.

    No.....1 Yes.....2

26. Your Social Security number: ___ ___ ___ - ___ ___ - ___ ___ ___ ___ (for payment purposes)

54020-16507 • Y99M.45 • Printed in the U.S.A.
I.N. 209217