College Board Report No. 99-4

The Performance Assessment Study in Writing: Analysis of the SAT® II: Writing Subject Test

HUNTER M. BRELAND, MELVIN Y. KUBOTA,
and MARILYN W. BONNER

The College Board

College Board Report No. 99-4
ETS RR No. 99-24

The Performance Assessment Study in Writing: Analysis of the SAT® II: Writing Subject Test

HUNTER M. BRELAND, MELVIN Y. KUBOTA, and MARILYN W. BONNER

College Entrance Examination Board, New York, 1999

Hunter M. Breland is senior research scientist at ETS.
Melvin Y. Kubota is assessment specialist II at ETS.
Marilyn W. Bonner is assessment specialist II at ETS.

Additional copies of this report (item #275900) may be obtained from College Board Publications, Box 886, New York, New York 10101-0886, (800) 323-7155. The price is $15. Please include $4 for postage and handling.

Printed in the United States of America.

Acknowledgments

In addition to the advisory committee, a number of people contributed to this project. The following persons assisted with data collections in the participating institutions: Melanie Levinson, Larry McCauley (The College of New Jersey); Brett Millier, Elizabeth Napier, Mark Williams (Middlebury College); Bronwyn Beistle, Janet Knepper, Aselda Thompson (University of Pennsylvania); Elizabeth Crighton, Eric Grosfils, Cristanne Miller (Pomona College); Debra L. Carney, Holly Davis, Mary A. Koncel (Smith College); Rachana Sachdev, Jeannie Zeck (Susquehanna University); Jennifer S. Lehman, Page Eileen Schilt (University of Texas–Austin); Kirsten Scheer (Ursinus College).

The following persons read and evaluated the writing samples: Lee Billmyer (Episcopal Academy, Marion, PA); Robert Brehme (Brookdale Community College, Lincroft, NJ); Casey Confoy (The College of New Jersey, Trenton, NJ); John Collins (Rowan University, Camden, NJ); Marilyn Collins (Rowan University, Camden, NJ); Hasel Cooper (Rowan University, Camden, NJ); Sally Ann Darnoi (Brookdale Community College, Lincroft, NJ); John Dawkins (Manor Junior College, Jenkintown, PA); Diane Gruenberg (New York University); Douglas Hoehn (Community College of Philadelphia); Robert Lorenzi (Camden Community College); Beatrice Moore (St. Joseph's University, Philadelphia, PA); RoseAnn Morgan (Middlesex Community College, Bedford, MA); Marti Palmere (The College of New Jersey, Trenton, NJ); Barbara Petrick (Dickinson High School, Jersey City, NJ); Linda Santora (J.F. Kennedy High School, Patterson, NJ); Irene Schiller (Camden Community College); Keith Slocum (Montclair State College, Upper Montclair, NJ); June Statmore (Wayne Township Public Schools, Wayne, NJ); C. David Wilhoite (Episcopal Academy, Merion, PA).

In addition to serving as readers, RoseAnn Morgan served as Chief Reader, and Diane Gruenberg and Robert Lorenzi served as Table Leaders. Thanks are also due to Matthew Tufts of ETS Performance Scoring Services, who made arrangements for the reading, and to Susan Taylor of ETS, who assisted with the reading.

Advisory Committee

Contents

Tables

Abstract

This study examined the SAT® II: Writing Subject Test as a predictor of writing performance in college English courses. Writing performance was based on eight writing samples submitted as part of regular course work by students in eight colleges. The samples consisted of drafts and final papers submitted in response to four take-home writing assignments. The samples submitted were evaluated by 20 experienced readers of student writing using a common scoring scale. Each sample was read and scored independently by two different readers. In addition to this writing performance criterion, two supplemental criteria were used: English course grades and students' self-assessments of their own writing ability and experience. In evaluating the SAT II: Writing Test as a predictor of these three criteria, special attention was given to comparisons of the predictive effectiveness of the essay and multiple-choice components of the test. It was concluded that both components were good predictors of the three criteria; however, the 40-minute multiple-choice component tended to produce higher predictive correlations than the 20-minute essay component. The best predictions were obtained when the two components were combined.

Key words: achievement tests, college performance, essay tests, higher education, predictive validity, writing tests

Introduction

The objective of this study was to examine the relationships between scores on the SAT II: Writing Subject Test and writing performance during the first year of college. The SAT II: Writing Test consists of two parts: (1) a 20-minute essay, which is scored independently by two different readers to arrive at a score, and (2) a 40-minute test with 60 multiple-choice questions on usage, sentence structure, and improving paragraphs. Subscores on both components are combined to determine the test taker's score on the College Board 200–800 scale. An important objective in the study was to compare the essay and multiple-choice parts of the test as predictors of college writing performance.

Because of the difficulty in conducting studies of this type, they tend to be relatively rare, and it is often necessary to go back quite a few years to find such studies in the literature. Huddleston (1954) studied 763 high school students who had written 20-minute essays and also taken a multiple-choice test of composition skill. Scores on the essays correlated .41 with high school English grades and .40 with teachers' ratings of the students' writing ability. The multiple-choice test of composition skill correlated .60 with English grades and .58 with teachers' ratings of the students' writing ability.

Godshalk, Swineford, and Coffman (1966) conducted a similar but more elaborate study. In this study, 24 public and private secondary schools collaborated. A total of 646 students each wrote five brief essays on the same topics, and each of the essays was read and scored independently by five different readers. The total score on the essays was then correlated with scores on several multiple-choice tests of composition skills. Each of the five essays was also correlated with composite scores for the other essays. The highest correlations with the essay score total were for multiple-choice tests of usage and sentence correction (both .71). The lowest correlation with the total essay score (.46) was for a multiple-choice test of paragraph organization. Single-essay scores (based on two readings) correlated .47 and .55 with a composite of the remaining four essays (each scored five times).

Breland and Gaynor (1979) studied over 2,000 students in four colleges who had taken a 20-minute essay pretest and a 30-minute multiple-choice pretest when applying for college. These students were administered essay posttests at the end of both the fall and spring semesters. The 20-minute essay pretest scores (based on two readers) correlated .52 with the fall essay posttest, .51 with the spring essay posttest, and .60 with the summation of the fall and spring posttests. The multiple-choice pretest correlated .58 with the fall essay posttest, .58 with the spring essay posttest, and .69 with the summation of the fall and spring posttests. The multiple-choice pretest correlated .72 with the summation of the fall essay pretest, the fall essay posttest, and the spring essay posttest.

Breland, Camp, Jones, Morris, and Rock (1987) collaborated with six institutions to obtain data for over 300 students taking freshman English composition courses. These students wrote six essays on different topics and in three different modes of discourse. Two of the six essays were take-home assignments. Each essay was read and scored independently by three different experienced readers. These students had also taken a 30-minute multiple-choice test of writing as a part of the college admission process. As in the Godshalk et al. (1966) study, a performance criterion was created by summing reader scores on the essays. Any single essay could then be evaluated by correlating its score with that of the other five essays. The average correlation of a single essay score (with two readings) with the five-essay criterion was .61. The average correlation of the multiple-choice test of writing with the same criterion was .62. When an essay was combined with the

multiple-choice test, an average multiple correlation of .72 was obtained.

The Breland et al. (1987) study also used course grades and instructors' ratings of student writing ability as criteria. The average correlation for a single essay read twice and course grades was .43. Scores on a 30-minute multiple-choice test of writing correlated .41 with course grades. When the essay and the multiple-choice test were used together, the multiple correlation with grades was .50. Instructors' ratings of student writing skill correlated .47 with a single essay, .48 with the multiple-choice test, and .56 when an essay and the multiple-choice test were used together.

Bridgeman (1991) conducted a study comparing essay and multiple-choice assessments of writing skill as predictors of grade-point average (GPA) in four state colleges and for a group of 21 colleges. For the four state colleges, a 20-minute essay test and a 35-minute multiple-choice test of sentence-level errors were used. Correlations of the essay score with GPA averaged .17 over 16 gender and ethnic groupings of students. The multiple-choice test of sentence-level errors correlated .26 on average for the same student groupings. For the grouping of 21 colleges, the multiple-choice and essay scores were from the English Composition Test (ECT, a forerunner to SAT II: Writing Test). The median of the correlation with GPA for the essay was .16 and, for a 40-minute multiple-choice test of writing, .30.

A preliminary investigation of the SAT II: Writing Test was conducted by Bridgeman and Bonner (1994). This study was conducted prior to the formal implementation of the SAT II: Writing Test by administering a preliminary version of the test as part of the regular placement-testing battery (in some colleges) and in freshman English composition courses (in other colleges). Scores on a 30-minute multiple-choice test of writing were also available for these students. A total of nine campuses and over 2,000 students were involved for the analyses comparing SAT II: Writing Test, SAT I verbal, and a 20-minute multiple-choice test of grammar and sentence structure. Most were state university campuses. Grades in regular freshman English courses were used as the criterion variable. Weighted average correlations with grades were .32 for the SAT II: Writing Test, .31 for the multiple-choice test of grammar and sentence structure, and .27 for SAT I verbal. No correlations were reported for the components of the SAT II: Writing Test (a 20-minute essay and a 40-minute multiple-choice section).

Powers, Fowles, and Boyles (1995) used grades in writing courses, students' self-assessments, accomplishments in writing, and reader scores on writing samples in a study of writing exercises. Median correlations of two 40- or 60-minute essays with criteria were as follows: .37 with grades, .39 with students' self-assessments, .30 with accomplishments in writing, and .27 with scored writing samples. When grades and accomplishments were combined, however, the median correlation increased to .53.

These previous studies suggest that both essay and multiple-choice tests of writing have reasonable relationships with criterion variables of importance. Studies that have examined both essay and multiple-choice tests indicate that the multiple-choice test usually yields higher correlations with criteria, but the highest correlations are obtained when essay and multiple-choice tests are used in combination.

The objective in the present study was to determine whether these same kinds of findings would hold for the SAT II: Writing Test introduced in 1994. The emphasis in the present study was on obtaining data on actual student performance in writing, as had been done in the previous studies by Godshalk et al. (1966) and Breland et al. (1987). Other data on course grades, students' self-assessments of writing ability, and student accomplishments in writing were also collected.

Data and Methods

Data were obtained from eight collaborating institutions for students entering college in fall 1996. All these institutions either required the SAT II: Writing Test for admission purposes or conducted special administrations of it early in the first semester of the first year of college. Each institution was requested to collect from each of approximately 40 students a total of four different writing samples from regular course work in first semester English composition courses. Each sample would have a draft version and a completed version. Although topics would be different in each institution, three general types of writing were requested: (1) response to text, (2) argument or persuasion, and (3) analysis. The four writing samples would consist of two samples of one type and one each of the other two types.

Student consent for participation in the study was obtained using the Consent Statement included in this report as Appendix A. Each student was also asked to complete a Writing Experience Questionnaire (Appendix B). This contained questions about students' grades in English courses, their own perceptions of their writing skills, and writing experiences including accomplishments such as winning an essay contest, being published, etc.

Institutions provided data on student SAT I verbal scores, SAT II: Writing Test scores (composite score plus subscores for the essay and multiple-choice components), and English composition course grades using the instruc-

tions and data sheets of Appendix C. Two institutions not requiring the SAT II: Writing Test for admission purposes conducted special administrations of it.

All data were coded by a student identification number created for the study to preserve anonymity.

Scoring of the writing samples was conducted at Educational Testing Service (ETS) by experienced readers. Although the usual scale used for most essay scoring at ETS has a 1 (low) to 6 (high) range, a special scoring scale was used for the present study. Because the samples of the present study were all from take-home assignments, they tended to be of higher quality than the usual timed essays. The rating scale used was similar to that used for the College-Level Examination Program® (CLEP®) English Composition examination. Exemplars of each point on the scoring scale were drawn from the papers collected for the study, and these exemplars thus became the guide for the score assignments. Writing samples labeled as drafts were scored first and final samples were scored second. Two readers read and scored each of eight samples for each student independently. Four topic scores were then developed as the simple sum of the two reader scores for the draft sample and the two reader scores for the final sample. These topic scores had a range from 8 to 32. These topic scores were then totalled for each student to create a Total Writing Performance variable with a range from 32 to 128. Because some students did not submit all eight writing samples, an Average Writing Performance variable was developed by adding the scores for all samples submitted and dividing by the number of samples. The Average Writing Performance variable was not computed for cases with less than four writing samples, however.

The Writing Experience Questionnaire was used to generate scores for overall GPA, Writing GPA, Writing Self-Assessment, and self-reported Writing Accomplishments. This questionnaire was adapted from a similar questionnaire used by Powers, Fowles, and Boyles (1995) and originally based on research conducted by Baird (1979) and extended by Stricker and Rock (1996). The GPA and Writing GPA scales ranged from 7(A) to 1 (F). The Writing Self-Assessment scale was developed by averaging the scores from Questionnaire Items 5 through 7, each having a rating of 5 (high) to 1 (low). The Writing Accomplishment scale was developed from Item 9 of the Questionnaire, and it was computed as the simple sum of the number of activities reported. Some activities were given double weight, however (e.g., Won an essay contest; Wrote poetry, fiction, or essays that were published; Wrote a play that was publicly performed; Wrote a screenplay for a film; Wrote a novel or full-length book). Since almost all students' scores on this variable were 10 or less (although a few were much higher), the scale was limited to a maximum of 10. These self-reporting assessment scales were standardized and then summed to form a Self-Report Composite score for which each of the self-assessments was equally weighted.

English course grades received from institutions were coded as follows: A (10), A− (9), B+ (8), B (7), B− (6), C+ (5), C (4), C− (3), D (2), and F (1).

A lower-bound estimate of the reliability of the Total Writing Performance score was made by means of coefficient Alpha for a composite score, where the composite score components were the four topic scores. Relationships between criterion variables and predictor variables were examined through correlation and regression analysis.

Results

Data available for analysis are described in Table 1. More cases were available for some variables than others, with the most cases available for English course grade (306) and substantial numbers of cases available for Average Writing Performance (294), Writing Accomplishment (301), Writing Self-Assessment (302), High School GPA (301), and High School Writing GPA (301). The last four variables came from the Writing Experience Questionnaire.

Only 222 students submitted all eight writing samples. Thus it was only for these students that the Total Writing Performance variable could be computed. The lower-bound estimate (coefficient Alpha) of the reliability of the Total Writing Performance composite score was .82. This estimate is slightly less than similar estimates made for composite writing performance criteria by Godshalk et al. (1966) and Breland et al. (1987), which were .84 and .88, respectively. One possible reason for a lower reliability estimate in the present study is that essays in different institutions were written on different topics, which made the task of scoring more difficult. A second reason for a lower estimate in the present study could be that only two readings of each essay were conducted, whereas Godshalk et al. used five readings per essay and Breland et al. used three readings per essay.

SAT II: Writing Test scores were not available for all cases because one institution primarily used scores from the multiple-choice part of the SAT II: Writing Test. Cases having the SAT I verbal score were reduced because two of the collaborating institutions did not require the test for admission. Table 1 also indicates the generally high ability level of the students in the study, with an average SAT I verbal score of 634 (as compared

TABLE 1

Variable Description

Variable	*Cases*	*Mean*	*S. D.*	*Range*
Background				
High School GPA	301	6.02	.89	3–7
High School Writing GPA	301	5.79	1.04	1–7
Writing Self-Assessment	302	3.65	.51	2–5
Writing Accomplishment	301	5.24	3.14	1–10
Test Scores				
SAT Verbal Score	222	634	83.8	300–800
SAT II: Writing Test Score	271	560	114	260–780
SAT II: Writing Test (Multiple Choice)	286	55.7	11.0	24–80
SAT II: Writing Test (Essay)	256	54.9	12.2	30–79
College Performance				
Total Writing Performance	222	84.6	13.7	45–113
Average Writing Performance	294	5.20	.92	2.70–7.06
English Course Grade	306	7.49	1.73	1–10

Notes: (1) Test scores for 1995–96 national samples of test-takers had means and standard deviations as follows:

Test Score	Mean	S. D.	Range
SAT Verbal	499	107	200–800
SAT II: Writing Test	555	99	200–800
Multiple Choice	55	10	20–80
Essay	55	10	20–80

(2) Correlations among test scores for national versus study samples:

	SAT-W	Multiple-Choice Score	Essay Score
SAT-W		.94	.78
Multiple-Choice Score	.94		.54
Essay Score	.73	.45	

(National sample below diagonal; study sample above diagonal)

to a national average of 499 for students taking the SAT I in 1995–96).

Table 2 shows gender differences in the sample, which included substantially more females than males. Females averaged better than males on self-reported Writing Accomplishment and college English Course Grade, but gender differences for other variables were not statistically significant. Previous studies have found gender differences

TABLE 2

Gender Differences

	Cases		*Mean*		*S. D.*		
Variable	*M*	*F*	*M*	*F*	*M*	*F*	d
Background							
High School GPA	124	177	5.96	6.07	.90	.90	.12
High School Writing GPA	124	177	5.77	5.94	1.30	1.10	.14
Writing Self-Assessment	125	177	3.69	3.61	.48	.53	−.16
Writing Accomplishment	126	175	4.83	5.54	3.29	3.00	.24*
Test Scores							
SAT Verbal Score	94	116	626	643	98.5	70.9	.20
SAT II Writing Test Score	101	159	560	555	110	117	−.05
SAT II: Writing Test (Multiple Choice)	111	164	55.6	55.4	10.5	11.2	−.02
SAT II: Writing Test (Essay)	94	153	55.7	53.9	11.0	11.6	−.16
College Performance							
Total Writing Performance	89	133	85.7	83.8	14.2	13.4	−.14
Average Writing Performance	119	172	5.13	5.18	1.01	.88	.05
English Course Grade	123	174	7.21	7.71	1.82	1.66	.29*

*$p < .05$.

favoring females in assessments of writing (Breland, 1977; Breland and Griswold, 1981; Breland and Jones, 1982; Breland et al., 1995; Bridgeman and Bonner, 1994; Bridgeman and McHale, 1996; Engelhard, Gordon, and Gabrielson, 1981; Golub-Smith et al., 1993; National Assessment of Educational Progress, 1994). The failure to observe gender differences in test score means for the present study is undoubtedly related to the samples available for analysis, which were small and nonrandom.

Predicting Writing Performance

Table 3 gives correlations between predictor variables (test scores and self-reports) and Writing Performance when all variables are included. The number of cases available for the SAT I verbal score considerably limits the analysis, however, because of the need to maintain a constant N for same sample correlations. Table 3 shows that the SAT I verbal score, the SAT II: Writing Test score, and the SAT II: Writing Test multiple-choice subscore all predicted Total Writing Performance reasonably well, with correlations ranging from .44 to .58, while the SAT II: Writing Test essay subscore correlation was substantially less (.21). For this table and subsequent tables, the statistical significance of differences in correlation coefficients was tested by two different methods. For differences in correlations obtained for the same sample, a t-test suggested by McNemar (1955, p. 148) was used. To test differences in correlations obtained from different samples (e.g., male or female), correlations were transformed to Fisher's Z. All tests for correlational differences were made at the .05 level, unless otherwise noted.

The correlations in Table 3 of .58, .48, and .44 between Total Writing Performance and SAT I verbal, SAT II: Writing Test, and the SAT II: Writing Test multiple-choice subscore, respectively, were all found to be statistically different from the correlation of .21 obtained for the SAT II: Writing Test essay subscore. These three highest correlates of Total Writing Performance are also significantly different than the correlations of .10, .27, .09, and .31 obtained for self-reports. Only the SAT I verbal score correlation of .58, however, is significantly different from the Writing GPA Self-Report correlation of .35. The SAT I verbal score correlation of .58 is also significantly different from the SAT II: Writing Test score correlation of .48 and the SAT II: Writing Test multiple-choice subscore correlation of .44. Note that the number of cases available for these analyses is somewhat restricted (N = 112).

A similar pattern of correlations was observed in Table 3 for Average Writing Performance, and statistical tests showed that, generally, differences were similar to those obtained for Total Writing Performance. The correlation of .43 obtained for the SAT II: Writing Test multiple-choice subscore is not statistically different from the correlation of .30 obtained for the SAT II: Writing Test essay subscore; however, the correlation of .54 obtained for the SAT I verbal score is statistically significant from the correlation of .43 obtained for the SAT II: Writing Test multiple-choice subscore.

The observed correlational differences for males and females in Table 3 in some cases appear to be large, but none of these differences attained statistical significance by two-tail tests because of the relatively small N's available. Nevertheless, the gender correlational difference for the prediction of Total Writing Performance from the SAT II: Writing Test essay subscore (.31 vs. −.02) represents a difference in Fisher's Z of .34, which is considered by Cohen (1988) to be a "medium" effect size and thus is worth noting. Moreover, this difference

TABLE 3

Predicting College Writing Performance From Test Scores and Student Self-Reports (Correlations)

Predictor Variables	*Total Writing Performance*			*Average Writing Performance*		
	Total Sample	*Females*	*Males*	*Total Sample*	*Females*	*Males*
Test Scores	N = 112	N = 69	N = 43	N = 154	N = 93	N = 61
SAT I Verbal Score	.58*	.64*	.49*	.54*	.58*	.58*
SAT II: Writing Test Score	.48*	.48*	.47*	.48*	.49*	.46*
SAT II: Writing Test Multiple-Choice Score	.44*	.39*	.52*	.43*	.40*	.46*
SAT II: Writing Test Essay Score	.21*	.31*	−.02	.30*	.37*	.20
Self-Reports						
High School GPA	.10	.25*	−.10	.25*	.28*	.21
High School Writing GPA	.35*	.39*	.26	.45*	.44*	.46*
Writing Self-Assessment	.27*	.34*	.11	.30*	.35*	.25*
Writing Accomplishments	.09	.16	.03	.19*	.18	.20
Self-Report Composite	.31*	.40*	.11	.42*	.43*	.42*

*$p < .05$.

is statistically significant by a one-tail test. The gender correlational difference for the prediction of Average Writing Performance represents a Z difference of .19, and is between Cohen's "small" effect size of .10 and a "medium" effect size. There is also a pattern in the gender correlations that is worth noting. While females appear to be better predicted by the SAT II: Writing Test essay subscore, males appear to be better predicted by the SAT II: Writing Test multiple-choice subscore for both writing performance criteria. Because of the pattern here, as well as in subsequent tables, the use of a one-tail test has some justification.

Table 4 gives correlations for predicting writing performance when the SAT I verbal score is excluded to allow for a larger number of data cases. The N is increased to 173 for Total Writing Performance and to 240 for Average Writing Performance. As was the case in Table 3, the correlations for the SAT II: Writing Test score (.56) and the SAT II: Writing Test multiple-choice subscore (.53) are not statistically significant, but the correlation for the SAT II: Writing Test essay subscore (.38) is significantly different from the other two test score correlations. And, as in Table 3, the correlations for the SAT II: Writing Test score and the SAT II: Writing Test multiple-choice subscore (.56 and .53, respectively) are significantly higher than that for the Self-Report Composite (.41). Similar observations can be made for predicting Average Writing Performance.

The gender differences in Table 4 do not reach statistical significance, even with the larger N. Nevertheless, the difference in the female and male correlations for predicting Total Writing Performance with the SAT II: Writing Test essay subscore (.43 and .24, respectively) represent a Fisher's Z difference of .22, which is between Cohen's (1988) "small" and "medium" effect size. The difference in the female and male correlations for the SAT II: Writing Test multiple-choice subscore (.48 and .63, respectively) represent a Fisher's Z difference of .22 as well. Similar effect sizes occur for the Average Writing Performance criterion.

Table 5 repeats the analyses of Table 4 after excluding Institution 8. This exclusion was suggested from observations that writing performance in Institution 8 was substantially lower than in the other institutions, even though test score performance was not lower. It was hypothesized that the writing topics used in Institution 8 were not as challenging as those used in other institutions and thus led to lower evaluations by the readers. Excluding Institution 8 increased most correlations substantially, but the pattern of differences is similar to that observed in Tables 3 and 4. The correlation for predicting Total Writing Performance using the SAT II: Writing Test score (.67) is significantly higher than that using the SAT II: Writing Test essay subscore (.52), and the difference between the correlation for the multiple-choice subscore (.61) and the essay subscore (.52) is significantly different at the .05 level. The correlations for both the SAT II: Writing Test score (.67) and the multiple-choice subscore (.61) are significantly higher than that for the Self-Report Composite (.45). A similar pattern was observed for the correlations predicting Average Writing Performance.

None of the correlational differences by gender in Table 5 was statistically significant at the .05 level, but the gender pattern of males better predicted by the multiple-choice score and females better predicted by the essay score persists with effect sizes in the "small" to "medium" range.

TABLE 4

Predicting College Writing Performance From Test Scores and Student Self-Reports (Correlations, With SAT I Verbal Score Excluded)

Predictor Variables	*Total Writing Performance*			*Average Writing Performance*		
	Total Sample	*Females*	*Males*	*Total Sample*	*Females*	*Males*
Test Scores	N = 173	N = 113	N = 60	N = 240	N = 150	N = 90
SAT II: Writing Test Score	.56*	.54*	.58*	.50*	.49*	.51*
SAT II: Writing Test Multiple-Choice Score	.53*	.48*	.63*	.47*	.43*	.53*
SAT II: Writing Test Essay Score	.38*	.43*	.24	.38*	.42*	.30
Self-Reports						
High School GPA	.26*	.32*	.13	.32*	.32*	.32*
High School Writing GPA	.37*	.35*	.40*	.42*	.38*	.48*
Writing Self-Assessment	.33*	.35*	.25*	.34*	.35*	.30*
Writing Accomplishments	.16*	.15	.18	.22*	.18	.27*
Self-Report Composite	.41*	.42*	.37*	.46*	.43*	.50*

*$p < .05$.

TABLE 5

Predicting College Writing Performance From Test Scores and Student Self-Reports (Correlations, With SAT I Verbal Score and Institution 8 Excluded)

	Total Writing Performance			*Average Writing Performance*		
Predictor Variables	*Total Sample*	*Females*	*Males*	*Total Sample*	*Females*	*Males*
Test Scores	N = 158	N = 103	N = 55	N = 203	N = 129	N = 76
SAT II: Writing Test Score	.67*	.64*	.69*	.64*	.63*	.61*
SAT II: Writing Test Multiple-Choice Score	.61*	.56*	.70*	.60*	.55*	.68*
SAT II: Writing Test Essay Score	.52*	.54*	.45*	.53*	.55*	.48*
Self-Reports						
High School GPA	.29*	.34*	.18	.33*	.35*	.30*
High School Writing GPA	.37*	.33*	.45*	.37*	.33*	.43*
Writing Self-Assessment	.37*	.35*	.37*	.39*	.38*	.38*
Writing Accomplishments	.22*	.18	.30*	.24*	.20*	.31*
Self-Report Composite	.45*	.42*	.51*	.47*	.44*	.52*

*$p < .05$.

Predicting College English Course Grades

Table 6 gives the results of an analysis predicting first-semester college English course grades from test scores. Because of different institutional grading standards, these analyses were conducted first within the institutions, corrected for restriction of range in test scores, and then averaged over institutions with weighting according to the number of cases available for analysis within each institution. The weighted average correlations in Table 6 for predicting college English course grades have a pattern similar to that observed for predicting writing performance in Tables 3, 4, and 5. The observed weighted average correlation for the SAT I verbal score (.52) is slightly higher than that for the SAT II: Writing Test score (.49), although this difference is not statistically significant. And, as for the predictions for writing performance, the SAT II: Writing Test score is a significantly better predictor for grades than is the SAT II: Writing Test essay subscore, with a correlation of .35. The difference in the weighted average correlation for the multiple-choice component (.37) is not significantly different from that for the essay component (.35).

Table 7 gives the results of a similar analysis for predicting first-semester college grades from student self-reports. This analysis was conducted separately from the test score analysis of Table 6 in order to maximize the number of cases in the analysis. As for the analyses predicting writing performance, the Self-Report Composite yielded a higher average weighted correlation with college grades than any of its components.

TABLE 6

Predicting First-Semester College English Course Grades From Test Scores (Correlations)

Institution	*N*	*SAT I Verbal*	*SAT II: Writing Test*	*SAT II: Writing Test Multiple Choice*	*SAT II: Writing Test Essay*
1	36	.69	.63	.55	.53
2	32	.53	.41	.29	.56
3	33	.42	.51	.25	.40
4	13	.40	.73	.36	.42
5	50	—	.23	.17	.25
6	12	.53	.57	.55	−.35
7	34	—	.65	.65	.22
8	37	.48	.50	.34	.41
Weighted Average	247	.52	.49	.37	.35

Note: Correlations have been corrected for restriction of range in test scores.

TABLE 7

Predicting First-Semester College English Course Grades From Student Self-Reports (Correlations)

Institution	*N*	*GPA*	*WGPA*	*WSA*	*WACT*	*SRC*
1	47	.14	.51	.44	.23	.51
2	36	−.05	−.20	.32	−.18	−.05
3	38	.18	.59	.36	.30	.49
4	13	.51	.06	−.23	.01	.10
5	50	.45	.31	.15	.04	.40
6	37	.43	.44	.12	.19	.40
7	33	.49	.60	.33	.20	.65
8	40	.64	.62	.49	.46	.68
Weighted Average	294	.33	.39	.28	.17	.42

Abbreviations: GPA = High School Grade-Point Average; WGPA = High School Writing Grade-Point Average; WSA = Writing Self-Assessment; WACT = Writing Accomplishment; SRC = Self-Report Composite.

Predicting the Self-Report Composite From Test Scores

A third way of evaluating the effectiveness of the test scores is to examine their relationships with the Self-Report Composite. The data from which the Self-Report Composite was developed were obtained from the Writing Experience Questionnaire administered toward the beginning of the first semester of college, although it is based on retrospective judgments made by students. Most of the test scores were obtained as part of the college admission process (except for two institutions that administered the SAT II: Writing Test as part of the study).

Table 8 shows correlations between test scores and the Self-Report Composite. When all test scores are included, the number of cases available for analysis is reduced to 156. For the total sample, only the difference in correlation between the SAT II: Writing Test score (.43) and the SAT II: Writing Test multiple-choice subscore (.32) is significantly different. When the SAT I verbal score is excluded, however, the number of cases available increases to 242, and the observed correlations follow a pattern similar to that obtained for previous analyses. The difference in the correlation between the SAT II: Writing Test score (.51) and the correlations for the multiple-choice subscore (.46) and the essay subscore (.43) are both statistically significant. The difference between the correlations for the multiple-choice component (.46) and the essay component (.43) is not statistically significant. None of the correlational differences by gender in Table 8 is statistically significant at the .05 level. The pattern of males being better predicted by the multiple-choice component observed in other analyses persists, with a Fisher's Z difference of .18, a "small" to "medium" effect size. The pattern of females being better predicted by the essay component did not occur, however, when the Self-Report Composite was used as the criterion.

TABLE 8

Predicting the Self-Report Composite From Test Scores (Correlations)

	Self-Report Composite		
Test Scores	*Total Sample*	*Females*	*Males*
Including All Test Scores	N = 156	N = 94	N = 62
SAT I Verbal Score	.38*	.31*	.48*
SAT II: Writing Test Score	.43*	.38*	.52*
SAT II: Writing Test Multiple-Choice Score	.32*	.26*	.42*
SAT II: Writing Test Essay Score	.39*	.38*	.41*
Excluding SAT I Verbal Score	N = 242	N = 151	N = 91
SAT II: Writing Test Score	.51*	.47*	.59*
SAT II: Writing Test Multiple-Choice Score	.46*	.41*	.55*
SAT II: Writing Test Essay Score	.43*	.42*	.46*

*$p < .05$.

Relative Contributions of Predictor Variables

The final issue examined was the relative contributions of various variables in the prediction of the criteria. This issue was examined through multiple regression analysis. Table 9 shows the results of analyses conducted when both the SAT I verbal score and the SAT II: Writing Test score are combined in predictions of Total Writing Performance, Average Writing Performance, and the Self-Report Composite. College grades were not used as a criterion because of different grading standards within institutions. Both the SAT I verbal score and the SAT II: Writing Test score contributed significantly in the prediction of Total Writing Performance, and the

TABLE 9

Predictive Contributions of SAT I Verbal and SAT II: Writing Test

Criterion Variable	*Predictors*	*N*	*r*	*beta*	*R*
Total Writing Performance	SAT I Verbal	127	.57	.50**	.59 (.58)
	SAT II: Writing Test		.50	.39*	
Average Writing Performance	SAT I Verbal	173	.52	.43**	.56 (.55)
	SAT II: Writing Test		.49	.40**	
Self-Report Composite	SAT I Verbal	169	.37	.24	.44 (.43)
	SAT II: Writing Test		.43	.32**	

*$p < .05$.
**$p < .01$.
Note: Figures for R in parentheses corrected for shrinkage.

multiple correlation of .59 is significantly higher than the zero-order correlation of .57 (.05 level). Similarly, in the prediction of Average Writing Performance, both predictors contributed significantly (at the .01 level), and the multiple correlation of .56 is significantly higher (.01 level) than the zero-order correlation of .52. In the prediction of the Self-Report Composite, only SAT II: Writing Test was a statistically significant contributor, but the multiple correlation of .44 was significantly higher (at the .01 level) than the zero-order correlation of .37. These multiple correlations represent increases in the squared multiple-correlation of between 2 percent and 5 percent, which are considered "small" to "medium" effect sizes by Cohen (1988).

Table 10 gives the results of analyses of the relative contribution of the SAT II: Writing Test subscores: the multiple-choice subscore and the essay subscore. In the prediction of Total Writing Performance and Average Writing Performance, the multiple-choice component received heavier weight, although both components contributed significantly in the predictions. Both components received roughly equivalent beta weights in the prediction of the Self-Report Composite. For all three criteria, the multiple correlations are significantly higher than the zero-order correlations for the multiple-choice component (.05 level), and the increases in multiple R represent increases in variance explained in the 2 to 6 percent range.

Discussion

Three different criteria were used to examine the effectiveness of the SAT II: Writing Test as a predictor of writing ability. The criteria were writing performance in college English courses, college English course grades, and students' self-reports of their writing ability and experience. Three different criteria were used because it was known that, because of the different meaning of grades in different institutions, college English course grades could be problematical. The writing performance criterion was emphasized in this study because of previous research indicating its superiority over other criteria (Godshalk et al., 1966; Breland et al., 1987). Self-reports by students were added as a third criterion type because of recent research indicating their utility (Powers, Fowles, and Boyles, 1995). The best predictors of writing performance were the SAT I verbal score and the SAT II: Writing Test score, with a composite of the self-report variables providing a reasonably good prediction. Generally, the multiple-choice component of the SAT II: Writing Test score yielded better predictions than did the essay component.

Because of differences in grading standards across institutions, the predictions of college English course grades were first made within each institution and corrected for restriction of range in test scores. Then a weighted average of the correlations was computed. The

TABLE 10

Predictive Contributions of SAT II: Writing Test Essay and Multiple-Choice Components

Criterion Variable	*Predictors*	*N*	*r*	*beta*	*R*
Total Writing Performance	Multiple-Choice Score	177	.53	.44**	.55 (.54)
	Essay Score		.39	.15*	
Average Writing Performance	Multiple-Choice Score	246	.47	.36**	.50 (.49)
	Essay Score		.38	.18**	
Self-Report Composite	Multiple-Choice Score	192	.42	.30**	.49 (.47)
	Essay Score		.38	.26**	

*$p < .05$.
**$p < .01$.
Note: Figures for R in parentheses corrected for shrinkage.

weighted mean correlations showed that the best predictors of English course grades were the SAT I verbal score, the SAT II: Writing Test score, and a composite of the self-reports. When the Self-Report Composite was used as a criterion variable, the SAT II: Writing Test score was a significantly better predictor than either of its components, as in the other predictive analyses.

The examination of the relative contributions of predictor variables was conducted by combining predictors in multiple regression analyses. These analyses showed that the SAT I verbal score and the SAT II: Writing Test score both made statistically significant contributions in the prediction of writing performance. In the prediction of the Self-Report Composite, however, only the SAT II: Writing Test score made a statistically significant contribution. A comparison of the multiple-choice and essay components of SAT II: Writing Test showed that both were statistically significant contributors in predictions of both writing performance and the Self-Report Composite. The most weight was assigned to the multiple-choice score in the prediction of writing performance, however. These multiple correlations also demonstrated that a combination of multiple-choice and essay components provides a better prediction than either component used alone. The use of SAT I verbal scores in addition to SAT II: Writing Test scores also improved the prediction of writing performance.

Such comparisons need to be made recognizing that the multiple-choice component is a 40-minute test with 60 questions and the essay component is only a 20-minute test with one question. If equivalent timing had been allowed for the essay component (with, say, two essay questions), the outcomes could have been different. The outcomes of this study are in general agreement with previous studies of this type. High correlations were obtained between predictor variables and criterion variables. This was especially the case for the writing performance criterion. And, as in two previous studies, an essay assessment added significantly to the prediction possible using a multiple-choice test alone. One difference between the present study and the previous studies is that some of the correlations observed were not quite as high as those observed in some studies. For example, Godshalk et al. (1966) and Breland et al. (1987) obtained correlations in excess of .70 between comprehensive performance assessments and multiple-choice measures of writing. The highest comparable correlation obtained in the present study was .61.

It is hypothesized that there are at least two reasons for this difference: The essay criterion used in both the Godshalk and Breland studies was concurrent, that is, the criterion data were obtained at the same time as the predictor data. In the present study, the predictor data were obtained when students were applying for admission to college, and the criterion data were obtained after students enrolled in college. The intervening time between the data collections would tend to attenuate correlations. A second difference between the present study and the two previous studies is that the essay criterion used in both the previous studies consisted of essays written on the same topics by all students. In the present study, students in different institutions submitted writing samples based on different writing assignments. Consequently, the scoring task was more complex and the estimated reliability of the composite writing performance criterion was lower, which would also tend to attenuate correlations.

There was an observed pattern of gender differences that was interesting but generally not statistically significant. It appeared that the college writing performance of females was better predicted by the essay subscore of the SAT II: Writing Test while males appeared to be better predicted by the multiple-choice subscore. Nevertheless, given that the usually observed female superiority in writing test scores was not observed in this sample of students, it does not seem appropriate to emphasize any observed gender differences.

It can be concluded that the SAT II: Writing Test is an effective predictor of writing performance in college. The 40-minute multiple-choice component of the test is a better predictor of writing performance in college than is the 20-minute essay component, but both combined into a single score provide the best prediction.

References

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

Breland, H. M. (1977). *Group comparisons for The Test of Standard Written English*. College Board Research and Development Report RDR 77-78, No. 1 (ETS Research Bulletin RB-77-15). Princeton, NJ: Educational Testing Service.

Breland, H. M., Bonner, M. W., & Kubota, M. Y. (1995). *Factors in performance on brief, impromptu essay examinations*. College Board Report No. 95-4 (Educational Testing Service RR No. 95-41). New York: College Entrance Examination Board.

Breland, H. M., Camp, R., Jones, R. J., Morris, M. M., & Rock, D. A. (1987). *Assessing Writing Skill*. College Board Research Monograph No. 11. New York: College Entrance Examination Board.

Breland, H. M., & Griswold, P. A. (1982). Use of a performance test as a criterion in a differential validity study. *Journal of Educational Psychology*, 74(5), 713–721.

Breland, H. M., & Jones, R. J. (1982). *Perceptions of writing skills*. College Board Report No. 82-4 (Educational Testing Service RR No. 82-47). Princeton, NJ: Educational Testing Service.

Bridgeman, B. (1991). Essays and multiple-choice tests as predictors of college freshman GPA. *Research in Higher Education*, 32(3), 319–331.

Bridgeman, B., & Bonner, M. W. (1994). SAT-W as a predictor of grades in freshman English composition courses. Unpublished report. Princeton, NJ: Educational Testing Service.

Bridgeman, B., & McHale, F. (1996). *Gender and ethnic group differences on the GMAT Analytical Writing Assessment*. Educational Testing Service RR No. 96-2. Princeton, NJ: Educational Testing Service.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*, second edition. Hillsdale, NJ: Erlbaum.

Engelhard, G., Gordon, B., & Gabrielson, S. (1991). The influences of mode of discourse experiential demand, and gender on the quality of student writing. *Research in the Teaching of English*, 26(3), 315–336.

Godshalk, F. I., Swineford, F., & Coffman, W. E. (1966). *The measurement of writing ability*. New York: College Entrance Examination Board.

Golub-Smith, M., Reese, C., & Steinhaus, K. (1993). *Topic and topic type comparability on the Test of Written English*. TOEFL Research Report 42. Princeton, NJ: Educational Testing Service.

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

McNemar, Q. (1955). *Psychological Statistics*. New York: John Wiley & Sons.

National Assessment of Educational Progress (1994). NAEP 1992 writing report card. Washington, DC: National Center for Education Statistics.

Powers, D. E., Fowles, M. E., & Boyles, K. (1995). *Validating a GRE writing test* (Graduate Record Examination Board Report No. 93-26B). Princeton, NJ: Educational Testing Service.

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

Appendix A: Student Consent Statement

Dear Student:

Your class is participating in a research study being sponsored by the College Board and conducted by Educational Testing Service (ETS). Your English department and several English departments nationally are participating. The purpose of the study is to evaluate writing tests used in college admission and placement. The results of the study will be used to improve these tests.

After removing your name and any other identification, your instructor will be sharing samples of your writing with ETS. These writing samples will be evaluated by writing experts working with ETS. The samples will be identified using an anonymous code that your instructor will assign and associate with writing test scores you received while applying for admission to college (or administered at your institution after you enrolled).

It is requested that you complete the enclosed questionnaire about your writing experiences. The purpose of the questionnaire is to obtain some indication of the amount of experience you have had with writing. The questionnaire responses will be associated with your writing samples and writing test scores through the same anonymous code assigned by your instructor.

Your course grade will not be affected in any way by the evaluations of your writing to be made by ETS. In fact, the ETS evaluations will not be conducted until after your course grade has been assigned. Nevertheless, if for any reason you do not want your writing samples shared with ETS and you do not want to participate in the study, please notify your instructor. There will be no penalty, implicit or otherwise, for not participating.

Appendix B: Writing Experience Questionnaire

1. What is your gender?
   a. Female
   b. Male

2. How do you describe yourself?
   a. African American/Black (non-Hispanic)
   b. American Indian/Native American/Alaskan Native
   c. Asian American/Pacific Island American
   d. Caucasian/White (non-Hispanic)
   e. Hispanic/Latino/Chicano/Mexican American/ Puerto Rican
   f. Other

3. Is English the dominant language in your household?
   a. Yes
   b. No

4. Do you understand English as well as or better than any other language?
   a. Yes
   b. No

5. How well do you think your writing compares with other students your age?
   a. Well above average 5
   b. Somewhat above average. 4
   c. About average. 3
   d. Somewhat below average 2
   e. Well below average 1

6. For each of the activities listed, please indicate how successful you were in high school writing assignments. (Circle one number for each.)

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

A. Thinking about an assignment (e.g., developing ideas, gathering information)	5	4	3	2	1
B. Organizing (e.g., making outlines, deciding on the order to present ideas)	5	4	3	2	1
C. Revising (e.g., improving sentence phrasing, rearranging ideas, correcting grammar and punctuation)	5	4	3	2	1
D. Developing an effective writing style (e.g., expressing ideas clearly and in interesting ways)	5	4	3	2	1

7. For each kind of writing listed, please indicate how successful you were in high school.

   5 = extremely successful 2 = not very successful
   4 = quite successful 1 = not successful at all
   3 = somewhat successful 0 = I did not do this

A. Personal writing experiences or feelings (as in a journal)	5	4	3	2	1	0
B. Creative writing (e.g., a poem or short story)	5	4	3	2	1	0
C. Persuasion (e.g., arguing a position or writing a letter to the editor)	5	4	3	2	1	0
D. Analysis or criticism (e.g., reviewing a book, movie, article, theory, work of art)	5	4	3	2	1	0
E. Description or explanation (e.g., describing an experiment or how something works)	5	4	3	2	1	0
F. Examination writing (e.g., long essay answers)	5	4	3	2	1	0
G. Applied writing (e.g., preparing an application or résumé)	5	4	3	2	1	0

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

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

9. Writing Activities

Descriptions of a variety of writing activities in school, out of school, in volunteer work, or in part-time or full-time jobs are listed below. Please read each description, and then indicate whether you have engaged in the activity by checking the "Yes" or "No" box next to the description. Some of the activities are relatively uncommon ones in which few people have been engaged.

Yes No

[ ] [ ] 1. Had a pen pal for less than a year.

[ ] [ ] 2. Had a pen pal for more than a year.

[ ] [ ] 3. Kept a nonschool journal or a diary for less than a year.

[ ] [ ] 4. Kept a nonschool journal or a diary for more than a year.

[ ] [ ] 5. Wrote a letter to the editor of a high school newspaper that was published.

[ ] [ ] 6. Wrote a letter to the editor of a town or city newspaper that was published.

[ ] [ ] 7. Wrote a letter to the editor of a magazine that was published.

Question 9 (continued):

Yes No

[ ] [ ] 8. Worked as a reporter for a high school newspaper.

[ ] [ ] 9. Worked on the editorial staff for a high school newspaper.

[ ] [ ] 10. Served as an assistant or associate editor of a high school newspaper.

[ ] [ ] 11. Served as the editor for a high school newspaper.

[ ] [ ] 12. Worked on the editorial staff for a high school yearbook.

[ ] [ ] 13. Served as an assistant or associate editor for a high school yearbook.

[ ] [ ] 14. Served as the editor for a high school yearbook.

[ ] [ ] 15. Worked as a reporter for a town or city newspaper.

[ ] [ ] 16. Participated in an essay contest.

[ ] [ ] 17. Won an essay contest.

[ ] [ ] 18. Wrote a feature article, column, or editorial that was published.

[ ] [ ] 19. Wrote a speech that you gave at a public gathering.

[ ] [ ] 20. Wrote a speech for someone else that was given at a public gathering.

[ ] [ ] 21. Wrote advertising or public relations material.

[ ] [ ] 22. Wrote technical manuals or other instructional material.

[ ] [ ] 23. Wrote poetry, fiction, or essays that were published.

[ ] [ ] 24. Wrote a play that was publicly performed.

[ ] [ ] 25. Wrote a screenplay for a film.

[ ] [ ] 26. Wrote a script for a dramatic or comedy show.

[ ] [ ] 27. Attended a writer's workshop.

[ ] [ ] 28. Wrote a short story that was not a part of schoolwork.

Question 9 (continued):

Yes	No	
[ ]	[ ]	29. Wrote a novel or other full-length book.
[ ]	[ ]	30. Other writing experience or activity: _____ _____ _____

Appendix C: Performance Assessment Study in Writing

Data Collection Instructions

Student Consent Statement

The Student Consent Statement is the one-page document that begins with "Dear Student." This statement has been reviewed and approved by the ETS Committee on Prior Review of Research. It informs the students that their institution is participating in a research study and assures them that the study will have no affect on their course grade or have any effect that might be perceived as harmful to them. It also assures students that they do not have to participate in the study and that there will be no penalty for not participating.

Copies of this statement should be made and distributed to students early in the semester or quarter.

Writing Experience Questionnaire

As explained in the Student Consent Statement, the purpose of the questionnaire is to obtain some indication of the amount of experience students have had with writing. It is especially important because it will represent the only information on students that will be comparable across participating institutions (the writing samples to be collected will be on different topics in each institution, and course grades are not comparable across institutions).

Copies of the Writing Experience Questionnaire should be distributed to students early in the semester or quarter. Students should be asked to return them to their instructor within a week. Institutional coordinators (Advisory Committee Members) should mail the completed questionnaires to ETS as soon as all students in their participating classes have completed them.

Writing Samples

The writing samples are the central focus of the study. They represent the actual performance of students in writing after they have enrolled in college and constitute the principal criterion through which the SAT II: Writing Test will be evaluated.

The Advisory Committee has agreed that eight samples of writing will be collected for each student as follows. Three modes or descriptive categories were agreed upon: (1) response to text, (2) argument or persuasion, and (3) analysis. For each of the three descriptive categories, two samples will be collected. The first sample in each category will be a draft essay and the second sample a finished essay. One additional draft and finished essay will be collected for any one of the three descriptive categories to make a total of eight writing samples for each student. It is important that all eight writing samples be collected for each student; any student without all eight will be excluded from the study.

The writing samples collected should be copied before instructor comments have been written on them. Additionally, student names must be removed and replaced with student identification codes (as explained below).

Institutional coordinators should return the writing samples to ETS in the Federal Express mailing envelopes provided as soon as they are collected, copied, and coded. A total of nine Federal Express mailing envelopes are provided in the study materials, eight for each of the eight sets of writing samples and one for the student roster.

Student Identification

Because students have been promised anonymity, all collected writing samples, questionnaires, and other information must be identified by student identification codes. These codes will be based on an initial digit representing the institution, a second digit representing the class section within the institution, and two additional digits to identify the student.

Student Rosters

Two types of student rosters will be required: (1) Student Roster #1, including student names, to be retained at the institution, and (2) a roster excluding student names to be returned to ETS. For clerical convenience, institutions may want to produce only a single original roster, copy it, and then clip off the student name column before returning it to ETS.

One page of each type of student roster is enclosed (copies can be made for use with different sections of students). Both rosters contain the following columns: Student ID, SAT I verbal score, SAT II: Writing Test score, SAT II: Writing Test (multiple choice), SAT II: Writing Test (essay), and course grade. The institutional roster (to be retained at the institution) includes a column for student name.

The Student Identification Code is created as described above. The test scores are to be obtained from institutional records or from a roster of students applying to your institution supplied by ETS. Because institutions enter the test score data, and because student names are to be removed from rosters before sending them to ETS, it will not be possible to associate student names with test

scores, writing samples, or other data. Course grade is to be obtained from institutional records, and it is requested that a description of the institutional grading system be attached to the roster returned to ETS.

It is important that all data be entered for all students. Any students with missing data will be excluded from the study.

Performance Assessment Study in Writing
Student Roster #1 (to be retained at institution)

Student Name	Student ID	SAT I Verbal Score	SAT II: Writing Test Score	SAT II: Writing Test (Multiple Choice)	SAT II: Writing Test (Essay)	Course Grade

Performance Assessment Study in Writing
Student Roster #2 (to be returned to ETS)

Institution _____

Student ID	SAT I Verbal Score	SAT II: Writing Test Score	SAT II: Writing Test (Multiple Choice)	SAT II: Writing Test (Essay)	Course Grade*

*Please attach information describing your institution's grading system.