

Comparing Essays Written Under Different Timing Conditions

Wayne J. Camara

In 1994 the College Board introduced the SAT[®] II: Writing Subject Test, which included a 20-minute timed essay and a 40-minute multiple-choice section. The time requirement for the essay was established after a review of research on the effects of essay timing on reader reliability, student scores, and task difficulty for students at different ability levels. Research from that effort is relevant today as we design the new SAT I writing section.

ETS conducted a study (Livingston, 1987) that examined differences in essay scores completed under three separate timing conditions: (a) 20 minutes, (b) 30 minutes, and (c) 30 minutes with separately timed sections of 10 minutes for planning and 20 minutes for writing. Two different essay topics were employed in the study.

Results showed that the two essays differed in difficulty—one essay was clearly easier for the majority of students irrespective of ability level, timing, or order of presentation. Several conclusions from that study were made concerning essay timing:

- The effect of an extra 10 minutes (allowing 30 minutes instead of 20 minutes) was “very small in relation to the other sources of variation,” and the effect of students’ ability on the difference between a 20- and 30-minute essay was far short of significant ($p = .23$).
- Providing students with 30 minutes, but requiring a 10-minute planning period (condition c above) appeared to lower scores if this essay came first, and appeared to have a slight increase in scores if it was the second essay.
- Additional analyses were conducted on students by ability level:
 - For students with low ability, neither the extra time nor topic made a difference in their score.

- Students who scored in the middle range of performance (scored between 6 and 8 on the 2–12 scale) were examined carefully because these are the students where course placement is most in doubt. Again, the difference between a 20-minute essay and 30-minute essay (in either conditions b or c) for this group did not even “approach statistical significance” (p. 10). The additional 10 minutes increased scores by 1/10 to 1/6 of a point on the 2–12 scale.
- For high-ability students, an extra 10 minutes (in either conditions b or c) appeared to improve scores by an average of 1/2 point.

Crone, Wright, and Baron (1993) also examined the effects of essay length in a study conducted to determine the final essay timing for the SAT II: Writing Test. Approximately 7,100 high school juniors and seniors completed several test sections from the SAT I verbal, SAT II: Writing Test (multiple-choice), the Test of Standard Written English, and two essays of 30 minutes or 15 minutes in length. Results clearly showed that examinees received lower essay scores in the 15-minute condition than in the 30-minute condition (the mean difference on a 2–12 scale was 1.22). Crone et al. (1993) determined that students were able to write reasonable essays in 15 minutes, albeit of marginally lower quality than 30 minutes, and with little impact on the overall rank order of students.

The study went further to examine if the time difference had any impact on ethnic/racial minorities or language minorities. The study confirmed that English Second Language (ESL) students¹ scored lower than English First Language (EFL) students irrespective of essay length and that all groups scored lower on the 15-minute essay.

1. Two separate questions were used to classify students into language groups. English first language and English best language were used to classify students into ESL and EBL groups. Both groups of students who report English is not their first language and not their best language are referred to as ESL for purposes of this report. The original study does provide separate analyses for ESL and EBL groups.

Research Summary

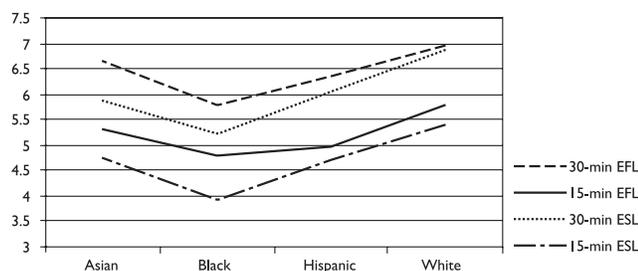


Figure 1. Average SAT-W essay scores for ethnic/racial groups by essay time and English First Language groups.

However, to determine whether any group is differentially disadvantaged by shorter essays, standardized differences should be computed. If ESL students are disadvantaged on the shorter essay, then the standardized differences between the ESL and non-ESL examinees would be larger for the 15-minute essay than for the 30-minute essay.

Figure 1 illustrates that there are no substantial or significant differences between the 15-minute and 30-minute essays for ESL and EFL students within any ethnic/race subgroup. Figure 2 shows that the standardized differences between ESL and EFL students are actually smaller with the 15-minute essay for three of the four groups.

Powers and Fowles (1996) examined the difference in examinee performance on a 40-minute and 60-minute proposed GRE writing test. Three hundred prospective graduate students completed two different essays under each of the time limits. On a questionnaire completed after writing the essays, 75 percent of respondents said a 40-minute time allocation was adequate, and 88 percent felt 60 minutes was adequate. The differences in the perception of time provided were statistically significant, especially for students who said they were slow or average test-takers. Additional time was equally benefi-

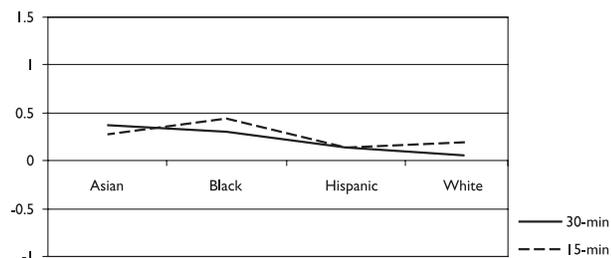


Figure 2. Standardized differences in SAT-W essay scores between EFL and ESL groups by ethnicity/race and essay time.

cial to test-takers who judged themselves as faster, average, or slower writers. Mean scores increased slightly with additional time (mean increases were .06 and 1.0 for different prompts on a 1–6 scale with two readers). However, the relative performance of fast, average, and slow test-takers and the meaning of test scores did not change noticeably when more time was allocated.

A similar study was conducted with the Test of English as a Foreign Language (TOEFL) in comparing the effects of 30-minute and 45-minute essays with 820 non-native English speakers (Hale, 1992). The correlation between scores for the 30-minute and 45-minute essays was .77, compared to .75 for essays written under the same time limit. The additional time increased scores by approximately 1/3 of a standard deviation but had little effect on the rank ordering of students. In addition, there was not a significant difference in the magnitude of the effect for students of low and high ability.

Walker (2002) conducted simulations using student data from the current SAT II: Writing Test to estimate the reliability of the new SAT I writing section under different timing conditions and weights for the multiple-choice (MC) and essay sections. He notes that “because each reader rates each essay globally on a 1-to-6 scale, and because this rating does not specifically take length into account, we can expect a similar distribution of scores and similar inter-rater agreement with a 25-minute essay as with a 20-minute essay” (p. 1). Whether the overall writing section is 50 minutes or 60 minutes, test reliability is higher with a shorter essay than with a longer essay because fewer objective items are included as essay length increases. A 60-minute writing section with a 20-minute essay would have the highest relative reliability and predictive validity, and a 50-minute writing section with a 25-minute essay would have the lowest relative reliability and predictive validity. The reason for this is that a 5-minute increase in essay time does not increase test reliability while a 5-minute increase in the MC section (about 8 additional items) does increase reliability.

Table 1 summarizes analyses of different test and essay lengths under standardized weighting.²

2. Standardized weighting allows the essay and multiple-choice (MC) sections to have equal standard deviations, as is done with the SAT II: Writing Test, and overweights the essay in relation to maximum scores. Other methods, such as unstandardized weighting, weight each section based on time allotted to it without any adjustments, and underweight the essay because the maximum score of 12 is so much less than the maximum score on the MC section. Finally, normalized weighting weights the essay and MC sections to have the same score range before weighting them by time allotted, with the weights for total score proportional to the testing time allotted to each section. Results for unstandardized and normalized weighting are similar to standardized weighting when examining the proportional differences between test timing (see Walker, 2002).

TABLE I
DIFFERENT TEST AND ESSAY LENGTHS
UNDER STANDARDIZED WEIGHTING

Essay time	60-minute test		50-minute test	
	20 min.	25 min.	20 min.	25 min.
Number of multiple-choice items	60	52	45	37
Multiple-choice test time	40 min.	35 min.	30 min.	25 min.
Test reliability	.87+	.84+	.84+	.79+
Scaled SEM	37	41	41	47
Validity with FGPA	.39	.38	.38	.37
Validity with college English grades	.42	.43	.42	.42

Results demonstrate that as reliability decreases, the scaled Standard Error of Measurement (SEM) increases. The SAT II: Writing Test currently has an SEM of 40 and reliability of .86 to .90 and is closest to the design in column 1.

In summary, previous research on differences in the reliability, validity, and difficulty of essay tests given under different timing conditions has indicated that giving examinees more time to complete an essay may raise their scores to a certain extent, but does not change the meaning of those scores, or the rank ordering of students. There is no evidence suggesting that giving less time to complete an essay advantages or disadvantages any particular ethnic or language subgroup, or ability level.

Wayne J. Camara is vice president of Research and Development.

REFERENCES

Crone, C., Wright, D., & Baron, P. (1993). *Performance of examinees for whom English is their second language on the spring 1992 SAT II: Writing Test* (Unpublished manuscript). Princeton, NJ: Educational Testing Service.

Hale, G.A. (1992). *Effect of amount of time allowed on the Test of Written English* (TOEFL Research Report #RR-92-27). Princeton, NJ: Educational Testing Service.

Livingston, S.A. (1987). *The effects of time limits on the quality of student-written essays*. Paper presented at the Annual Meeting of the American Educational Research Association, New York, NY.

Powers, D.E., & Fowles, M.E. (1996). Effects of applying different time limits to a proposed GRE writing test. *Journal of Educational Measurement*, 33 (4), 433-452.

Walker, M. (2002). *Properties of a writing test with a 25-minute essay*. Personal correspondence.

Office of Research and Development
 The College Board
 45 Columbus Avenue
 New York, NY 10023-6992
 212 713-8088

Copyright © 2003 by College Entrance Examination Board. All rights reserved. College Board, SAT, and the acorn logo are registered trademarks of the College Entrance Examination Board. Other products and services may be trademarks of their respective owners. Visit College Board on the Web: www.collegeboard.com.