RESEARCH REPORT 2012-9 (REV: 4-2013)

CollegeBoard

The SAT® Essay and College Performance: Understanding What Essay Scores Add to HSGPA and SAT

By Emily J. Shaw and Jennifer L. Kobrin

Emily J. Shaw is an associate research scientist at the College Board.

Jennifer L. Kobrin is a research scientist at the College Board.

Acknowledgments

The authors would like to thank Ed Hardin, Ying Liang, and Sunny Niu for their helpful feedback on previous versions of this manuscript.

About the College Board

The College Board is a mission-driven not-for-profit organization that connects students to college success and opportunity. Founded in 1900, the College Board was created to expand access to higher education. Today, the membership association is made up of over 6,000 of the world's leading educational institutions and is dedicated to promoting excellence and equity in education. Each year, the College Board helps more than seven million students prepare for a successful transition to college through programs and services in college readiness and college success — including the SAT® and the Advanced Placement Program®. The organization also serves the education community through research and advocacy on behalf of students, educators and schools.

For further information, visit www.collegeboard.org.

The SAT®

Created and designed by educators, the SAT® is a valuable and reliable measure of college readiness for students seeking admission to undergraduate colleges and universities in the United States. The SAT tests the academic skills and knowledge that students acquire in high school. It also shows how well students can apply their knowledge, a factor that educators and researchers agree is critical to success in college course work. The SAT is consistently shown to be a fair and valid predictor of first-year college success for all students. A study including data from more than 100 colleges and universities demonstrates that the best predictor of college success is a combination of SAT scores and high school grades. In addition to admission, many colleges use the SAT for course placement. The SAT is administered annually to more than two million students at approximately 6,000 test centers located in more than 170 countries.

VALIDITY

For more information on College Board research and data, visit research.collegeboard.org.

Contents

Executive Summary

This study examines the relationship between students' SAT® essay scores and college outcomes, including first-year grade point average (FYGPA) and first-year English course grade average (FY EngGPA), overall and by various demographic and academic performance subgroups. Results showed that the SAT essay score has a positive relationship with both FYGPA and FY EngGPA. Additionally, the SAT essay score offers useful and unique information in understanding how students perform in college generally, and in their English course work, after controlling for relevant academic variables (e.g., high school grades and SAT scores). As we continually seek to improve students' writing performance in the United States, the findings warrant further exploration of the value and role of the SAT essay in college admission and placement, particularly because students who take the SAT and apply to college are already providing this information to admission offices.

Introduction

The SAT writing section, which was added to the SAT in March 2005, includes both multiple-choice items and an essay. The multiple-choice items assess the ability to identify sentence errors and improve grammar and usage, and students have 35 minutes to answer 49 multiple-choice questions. The essay tests a student's ability to write effectively in response to an issue provided in a prompt, and students have 25 minutes to formulate their response. The SAT writing section is scored on a 200–800 scale, with the multiple-choice items accounting for approximately 70% of a student's writing score and the essay accounting for the remaining 30%. The essay is scored holistically by two independent readers on a scale of 1 to 6. If the two readers' scores differ by more than one point, a third reader scores the essay. Approximately 1% of all scored essays require a third reader (College Board, n.d.). The combined score for both readers ranges from 2 to 12, and essays that are not written on the essay assignment receive a score of 0. Official SAT score reports that are sent on behalf of applicants to colleges and universities include the essay subscore.

There has been a great deal of validity research on the SAT writing section as a whole since its introduction in 2005 (e.g., Kobrin, Patterson, Shaw, Mattern, & Barbuti, 2008; Mattern, Patterson, Shaw, Kobrin, & Barbuti, 2008; Patterson, Mattern, & Kobrin, 2009; Patterson & Mattern, 2010); however, there has been far less research focused on the essay component alone, and particularly on the relationship between the SAT essay score and college outcomes. Early research on the SAT essay (e.g., Kobrin, Deng, & Shaw, 2007) focused on examining data to study media criticisms of the SAT which claimed, for example, that the essay score merely reflected the number of words the student wrote or the nature of the examples provided. Kobrin, Deng, and Shaw (2011) examined the relationship between features of student responses on the essays and the scores they received. Their results indicated that essay length was related to scores, but the correlation was not nearly as high as previous critics have claimed. After controlling for SAT critical reading and writing multiple-choice scores, the essay features with the largest positive effect sizes included, using a five-paragraph theme and using academic evidence, although these effect sizes are still considered small. The features with the largest negative effect sizes included using no evidence or support and ending the essay mid-sentence.

Many educators have advocated for essay assessments (also often referred to as direct writing assessments) over indirect writing assessments such as multiple-choice questions. Direct writing assessments are viewed as more authentic and as having more positive "washback" effects than indirect writing assessments. These washback effects are similar to the notion of consequential validity, pertaining to the social consequences of using a particular test for a particular purpose (Messick, 1989; Noeth & Kobrin, 2007). That is, arguments have been made for more essay testing because traditional multiple-choice measures of writing skills are viewed as having the unintended side effect, or negative consequence, of causing teachers to teach and students to focus on sentence-level problems while ignoring the more global aspects of writing (Breland, Bridgeman, & Fowles, 1999).

However, the greater authenticity of direct writing assessment is usually offset by the lower reliability of scores compared to indirect writing assessments. Essay assessments have lower reliability than multiple-choice tests due to issues related to the number and selection of tasks, and issues related to scoring. When the predictive value of a test is measured by the correlation of students' scores on the test with college outcome measures such as first-year grade point average and grades in college courses, the indirect writing assessments will always prevail.

Nevertheless, research has shown the value of essay tests as predictors of college performance, especially when they are used in conjunction with multiple-choice writing test scores. For example, Breland (1983) cited five studies providing useful evidence for the incremental validity of direct writing assessments (essay tests) above high school GPA or rank, SAT scores, and other measures. In general, studies of writing tests for predicting college outcomes have suggested that both indirect and direct writing tests (essays) may be a slightly better predictor of English course grades than of more general outcomes such as GPA or instructors' ratings as criteria. Bridgeman and Lewis (1991) found that direct writing tests had greater predictive value in courses that were more likely to use essays in the determination of course grades, such as English or history courses rather than other outcomes.

Examining writing performance as the outcome, Breland, Kubota, and Bonner (1999) studied the predictive validity of the SAT Subject Test in writing (the precursor to the SAT writing section) for predicting writing performance in college English courses for students entering eight different colleges in the fall of 1996. College writing performance was based on eight writing samples (drafts and final papers submitted in response to four take-home writing assignments) submitted as part of regular course work by approximately 300 students. The SAT Subject Test writing scores were also used to predict students' self-reported writing ability. The multiple-choice and essay components were both statistically significant predictors of both English course writing performance and students' self-reports; however, the multiple-choice score was a stronger predictor of writing performance. The authors noted that a combination of the multiple-choice and essay components provided a better prediction than either component used alone.

Similarly, Matzen and Hoyt (2004) found that a timed essay test, in addition to a multiple choice writing assessment, was most valuable for writing course placement at an open admission, four-year public college in the Western United States. Students were misplaced into writing courses when relying solely on multiple-choice writing scores (about 30% were misplaced into lower level courses and about 9% were misplaced into higher level courses). Contrary to many other studies comparing direct and indirect writing assessments, Matzen and Hoyt found that if only one single measure could be used for placement, the timed essays were more accurate in placing students into English courses than the more indirect measures of writing.

Subgroup differences on both direct and indirect writing tests are a persistent concern in measurement. Breland et al. (1999) reported that women tend to perform better than men do on both types of writing assessments (Willingham & Cole, 1997). Minority ethnic groups and groups for whom English is not their best language perform relatively less well on indirect tests of writing than they do on direct writing assessments (Pomplun, Wright, Oleka, & Sudlow, 1992). To the extent that essay assessments are less reliable than multiple-choice tests, they would be expected to show smaller group differences.

The purpose of this study was to more deeply understand SAT essay performance by various subgroups and demonstrate the utility of the SAT essay, a direct writing assessment, as a predictor of college performance across students. A graphical or straightforward approach is used in this study to show what SAT essay scores add above HSGPA and composite (added together) SAT critical reading, math, and writing (multiple choice only) scores to the prediction of first-year college grades and first-year English course grades.

Data and Variables

Sample

The sample in this study was taken from the 110 four-year institutions that participated in the national SAT Validity Study and provided first-year performance data on the first-year, first-time students that entered college in the fall of 2007 (for more information on the sample, see Patterson et al., 2009). This sample of students (*N* = 216,081) was matched to the 2007 College-Bound Seniors database, which included the SAT scores and SAT Questionnaire responses of students who reported that they would graduate from high school in 2007. In order to be included in the study sample, students had to have valid SAT scores, first-year grade point average (FYGPA), first-year English grade point average (FY EngGPA), retention to the second-year information, and a self-reported high school grade point average (HSGPA) from the SAT Questionnaire, resulting in a final sample size of 120,897 students.

Measures

Demographic information. Self-reported gender, ethnicity, best language spoken, first language spoken, and highest parental educational level were obtained from student responses to the SAT Questionnaire, which is completed during registration for the SAT.

High school grade point average (HSGPA). Self-reported HSGPA was obtained from the SAT Questionnaire. This variable is on a 12-point scale with the following values: A+ (97–100; 4.33); A (93–96; 4.00); A– (90–92; 3.67); B+ (87–89; 3.33); B (83–86; 3.00); B– (80–82; 2.67); C+ (77–79; 2.33); C (73–76; 2.00); C– (70–72; 1.67); D+ (67–69; 1.33); D (65–66; 1.00); E or F (Below 65; 0.00).

First-year grade point average (FYGPA). Each participating institution supplied FYGPA values for the first-time, first-year students who entered in fall 2007. The range of FYGPAs across institutions was from 0.00 to 4.19.

First-year English grade point average (FY EngGPA). Each participating institution supplied all first-year course grades for the entering class of fall 2007. Each course was coded for its specific content area for analysis and all English course grades that a student took in the first year were averaged to arrive at the FY EngGPA. When Shaw and Patterson (2010) examined first-year college course work in English across a national sample of four-year institutions, they found that the vast majority of English courses taken were composition courses (72%). This is likely because most institutions have writing course requirements in the first-year. For the current study, when a student took one English course in the first year, this one course grade served as the FY EngGPA. The range of FY EngGPA across institutions was from 0.00 to 4.33.

Retention to the second year. In the fall of 2008, each participating institution supplied second-year retention information for students in the entering class of fall 2007. The retention indicator was dichotomously coded to indicate either "yes, the student returned" or "no, the student did not return."

SAT scores. Official SAT scores were obtained from the 2007 College-Bound Seniors database for the analyses. This database is comprised of the students who participated in the SAT program and reported that they would graduate from high school in 2007. A student's most recent score was used in the analyses. The SAT is comprised of three sections: critical reading, math, and writing, and each section is scored on a 200–800 scale The SAT writing section has two separate scores for the essay and the multiple-choice items, and these scores are combined to form a total SAT writing score.

Method

The analyses in this study included the calculation of means, standard deviations, and effect sizes or standardized differences using Cohen's *d* (1988) for SAT essay scores by subgroups. In this study, Cohen's *d* is calculated by subtracting the total group SAT essay score mean from the subgroup SAT essay score mean and dividing that resulting value by the total group's SAT essay score standard deviation. In order to interpret Cohen's *d*, Cohen provided basic guidelines, characterizing an effect size of .2 as small, an effect size of .5 as medium, and an effect size of .8 as large. However, Cohen emphasized that when interpreting effect sizes, the researcher must contextualize the data and use personal judgment to assess the practical significance of an effect.

Also, a graphical approach was incorporated to show what SAT essay scores add above HSGPA and composite SAT critical reading, math and writing multiple-choice scores in relation to college outcomes. This approach was advanced by Bridgeman, Pollack, and Burton (2004), and is often referred to as the *straightforward approach* because of its intuitiveness and simplicity as compared to the more traditional approach of showing incremental validity using multiple regression. As Bridgeman et al. noted, it is important to find ways to present the complicated concept of "variance explained" by different predictors in more straightforward ways for deeper understanding among broader audiences. In the current study, the academic performance groupings of High Achievers, Moderately High Achievers, Average Achievers, and Below Average Achievers (based on students' SAT critical reading, math, and writing multiple-choice scores and HSGPA) were based on those used by Bridgman et al., after confirming their face validity with regard to student performance standards.

Results

Descriptive Statistics

Means, standard deviations, and effect sizes for essay scores by student subgroups are included in Table 1. The overall mean essay score for the sample was 7.82 (SD = 1.49), indicating that this sample is more highly able than the typical population of college-bound seniors. In 2008, the first year that the essay subscores were reported in the annual *College-Bound Seniors* reports, the overall mean essay score was 7.10 with a standard deviation of 1.70. Consistent with trends found in the population, the female students in the study sample tended to have higher mean essay scores (M = 7.88, SD = 1.41, d = 0.04) than male students (M = 7.73, SD = 1.57, d = -0.06); however, the difference was quite small. With regard to racial/ethnic differences in mean essay score, Asian students had the highest mean scores (M = 8.13, SD = 1.56, d = 0.21), while black/African American students had the lowest mean scores (M = 7.15, SD = 1.44, d = -0.45). Students whose parents had received no high school diploma had a mean essay score of 7.16 (SD = 1.48, d = -0.45) and those with a parent who attended graduate school had a mean essay score of 8.21 (SD = 1.47, d = 0.26).

Mean essay score was also reported for best language and first language spoken subgroups. Interestingly, these variables followed different trends from each other. There were larger differences in mean essay scores for the best language subgroups, with the English as Best Language subgroup having the highest mean (M = 7.84, SD = 1.48, d = 0.01), followed by the English and Another Language subgroup (M = 7.63, SD = 1.56, d = -0.13), and the Another Language subgroup (M = 7.06, SD = 1.63, d = -0.51). For the first language spoken subgroups, those in the English and Another Language subgroup had the highest mean essay score (M = 7.86, SD = 1.56, d = 0.03), followed by those in the English as Best Language subgroup

(*M* = 7.82, *SD* = 1.47, *d* = 0.00), and those in the Another Language subgroup (*M* = 7.71, *SD* = 1.60, *d* = -0.07).

It is useful to note that many of the standardized differences on the SAT essay by subgroup can be compared to standardized differences by subgroup in earlier research on the SAT critical reading (or SAT Verbal in research on the test administered before March 2005) and math sections. In particular, Kobrin, Sathy, and Shaw (2007) presented such subgroup differences on the SAT and other academic measures over a period of 20 years. Examining the differences in magnitude of the standardized differences on the essay by subgroup as compared to the other SAT sections, it is apparent that there tend to be smaller standardized differences on the SAT essay than the SAT critical reading (and math sections), although there were some exceptions. The standardized differences in SAT critical reading scores for Hispanics students ranged from -0.49 in 2003 to -0.41 in 2006, while the standardized difference on the SAT essay score for Hispanic students in this study was -0.26 in this study. For black/African American students, the standardized differences in SAT critical reading scores ranged from -0.72 in 1987 to -0.61 in 2006. However, the standardized difference in SAT essay score for black/African American students in this study was -0.45. An exception is that the standardized difference on the essay for Asian students in this study was 0.21; however, the standardized differences on SAT critical reading scores in Kobrin et al. tended to be somewhat smaller for this subgroup (ranging from -0.08 in 1996 to 0.06 in 2006), while the difference was typically much larger for SAT math (ranging from 0.45 in 1996 to 0.52 in 2006). For students reporting their best language to be a language other than English, the standardized difference on the SAT essay (-0.51) was much smaller than that of SAT critical reading in the Kobrin et al. study (ranging from -1.05 in 1996 to -0.79 in 2006) but much larger than that of SAT math (ranging from 0.05 in 1996 to 0.12 in 2006), which requires less language proficiency.

Table 2 shows the relationship between SAT essay score and academic performance subgroups. There were rather large differences in mean essay score by SAT performance subgroups, which were created based on students' composite or sum of SAT critical reading, math, and writing (multiple choice only) scores. Mean essay score followed the trend of increasing SAT performance with the strongest SAT performance subgroup having a mean essay score of 9.56 (*SD* = 1.34, *d* = 1.17) and the weakest SAT performance subgroup having a mean essay score of 6.04 (*SD* = 1.42, *d* = -1.19). Differences in mean essay score by HSGPA groups were not quite as pronounced as in the SAT performance subgroups, though the patterns of essay performance were the same. Students with a HSGPA above an A- had a mean SAT essay score of 8.06 (*SD* = 1.45, *d* = 0.16), whereas students with a HSGPA below B- had a mean SAT essay score of 6.82 (*SD* = 1.49, *d* = -0.67). Mean SAT essay performance was also examined by college performance subgroups, including FYGPA, FY EngGPA, and retention to the second year. Students with a FYGPA above a B+ had a mean SAT essay score of 8.28 (*SD* = 1.43, *d* = 0.31), while students in the lowest FYGPA subgroup of Below D+ had a mean SAT essay score of 6.79 (*SD* = 1.48, *d* = -0.69). For FY EngGPA, subgroups showed similar trends in mean SAT essay score to FYGPA subgroups, although the magnitude of the differences was smaller than for FYGPA. The group with a FY EngGPA above B+ had a mean SAT essay score of 8.10 (*SD* = 1.43, *d* = 0.19), while students in the lowest FY EngGPA subgroup of Below D+ had a mean SAT essay score of 6.97 (*SD* = 1.52, *d* = -0.57). With regard to retention to the second year, those students who did return had a mean SAT essay score of 7.87 (*SD* = 1.48, *d* = 0.03) versus a mean SAT essay score of 7.37 (*SD* = 1.50, *d* = -0.30) for students who did not return.

Figure 1 shows the relationship between SAT essay score and mean FYGPA (black line) and mean FY EngGPA (blue dotted line). This figure shows that as essay scores increase, both GPAs also increase, and the mean FY EngGPA tends to be slightly higher than the mean FYGPA at each essay score point, except for those receiving the lowest essay scores.

Straightforward Approach

Because the essay score scale is much smaller, and scores are less variable than scores on the other SAT sections, using a traditional regression approach to determine the contribution of essay scores to the prediction of college grades would most likely lead one to incorrectly surmise that essay scores do not provide any increment to the prediction. As will be shown in our analysis, this conclusion would be misinformed. The *straightforward approach*, or graphical presentation of the data, depicts a different story.

Figure 2 shows the clear relationship between SAT essay scores and FYGPA, after essentially controlling for SAT scores and HSGPA. Table 3 includes the values that Figure 2 is based on. The table and figure group students by different achievement levels, based on their HSGPA and SAT critical reading + math + writing multiple choice scores, into: High Achievers, Moderately High Achievers, Average Achievers, and Below Average Achievers. While the trends are clear in the table, the graph more clearly shows the relationship of interest. With few exceptions (primarily due to smaller sample sizes), one can see that within all groups of similarly academically able students, as essay scores increase, so too does the mean FYGPA. The standardized differences in Table 3 show that within achievement levels, there are sometimes sizable differences in FYGPA by SAT essay score. For example, among Average Achievers, those with an essay score of 4 had a mean FYGPA of 2.23 (SD = 0.79, d = -0.37), whereas those with an essay score of 10 had a mean FYGPA of 2.69 (SD = 0.60, d = 0.28).

Not surprisingly, there were too few or no students with essay scores of 9, 10, 11, or 12 in the Below Average Achievers group to report on, just as there were too few High Achievers with an essay score of 2 or 3 to report on. Also, within each essay score point (e.g., all students with a score of 8, for example), trends across achievement level groups remained consistent with mean FYGPA, increasing steadily by achievement level.

Figure 3 depicts the relationship between SAT essay scores and FY EngGPA, while grouping students by SAT scores and HSGPA, or controlling for academic ability. Table 4 includes the values that Figure 3 is based on. Similar to the FYGPA relationship, within all groups of similarly academically able students, as essay scores increase, so too does the mean FY EngGPA. The standardized differences in Table 4 indicate that within achievement levels, there are sometimes sizable differences in FY EngGPA by SAT essay score. The differences in FY EngGPA by SAT essay score appear to be most pronounced for the Average Achievers group. For example, among Average Achievers, those with an essay score of 4 had a mean FY EngGPA of 2.31 (SD = 1.08, d = -0.39), whereas those with an essay score of 10 had a mean FY EngGPA of 2.96 (SD = 0.75, d = 0.30). Also, within essay score, trends remained consistent across achievement level groups, with mean FY EngGPA increasing steadily by achievement level. The few disrupted trend patterns appear to be due to small sample size.

Discussion

The results of this study show that the SAT essay has a positive relationship with both FYGPA and FY EngGPA. This relationship remains apparent even when students are grouped by academic ability (HSGPA and SAT composite score, not including the essay), providing evidence for the predictive validity of the SAT essay with FYGPA and FY EngGPA. In other

words, when HSGPA and SAT scores are held constant, the essay still provides important information regarding college performance.

With regard to subgroup differences on the SAT essay, consistent with prior research (Breland et al., 1999), female students tended to have higher essay scores than male students. Interestingly, however, Asian students in this sample outperformed other ethnic groups on the SAT essay and, with the exception of white students, outperformed other ethnic groups on the total SAT writing section in the 2007 College-Bound Seniors report (d = 0.17). This is unlike the historic performance of Asian students on the SAT Subject Test in writing (multiple choice and essay), the precursor to the SAT writing section on which Asian students performed about .3 of a standard deviation lower than the total group (Kobrin et al., 2007). This performance difference may be due to differences in the SAT Subject Test sample, as the students who took the SAT Subject Test in writing tended to be academically stronger students than general SAT takers because the SAT Subject Test in writing was used in admission to the University of California and other prestigious institutions.

There are clear differences in essay performance when the question of whether a language other than English is a student's first language or best language is considered. There were larger differences in SAT essay means when the students' self-reported best language was considered, as opposed to first language. As best language is likely indicative of students' self-efficacy beliefs in the English domain (Bandura, 1993, 1997), it is not surprising that this is highly related to performance on the SAT essay. Additionally, SAT essay means increased as all academic measures increased (e.g., the FYGPA and HSGPA subgroups), establishing a clear, positive relationship between the SAT essay and academic performance in this study, including the more noncognitively based retention variable.

The *straightforward approach* (Bridgeman et al., 2004) was used in this study because of the way it accessibly shows relationships between variables, requiring only an interpretation of trends in group means as opposed to more complicated statistics. Mattern, Kobrin, Patterson, Shaw, and Camara (2009) believe that this approach to explaining and showing relationships between variables such as the SAT and FYGPA may result in a better understanding and decreased misconceptions of research findings because the traditional multiple regression results providing measures of the "variance accounted for" are widely misinterpreted. By examining the relationship between SAT essay score and FYGPA or FY EngGPA, when students are grouped by academic ability, one can test whether the SAT essay is of additional assistance in predicting college performance above traditional measures such as HSGPA and SAT section scores. If the SAT essay did not help predict FYGPA or FY EngGPA, above these measures, one would see a straight line across all High Achievers (same mean FYGPA or mean FY EngGPA) and a straight line below for all Average Achievers, for example. This was not, however, the case. Students at each of the academic ability levels showed increases in mean FYGPA or FY EngGPA with increases in SAT essay score, except in a few cases, where it appears that the sample was quite small and therefore less stable (at the lower essay scores in this college-going sample). As a practical example, among all those students considered to be High Achievers, those with the lowest essay score that could be reported (4), had a FYGPA of 2.94, while those with the highest essay score (12) had a mean FYGPA of 3.50. This shows that by further grouping these high achieving students based on their performance on a 25-minute direct writing assessment, one will find differences in how these students will perform in college — evident in both their FYGPA and also specifically in their first-year English course work performance.

There are a few limitations of the study that should be reported. As the study sample included somewhat higher-achieving, college-going students, there were fewer students with essay

scores at the lower end of the scale, which limits the generalizability and conclusions that can be drawn at that part of the essay score scale. Also, the academic ability categories used in the study were created based on those used in Bridgeman et al. (2004) after confirming their face validity with regard to student performance standards. Although we wouldn't expect the patterns in the straightforward approach graphs to considerably change if the academic ability categories were altered, it would be useful to confirm this in future research.

Future research should explore the role of the SAT essay in college English or writing course placement decisions, given the relationship shown between essay score and English performance averaged across courses. It would also be useful to explore how the relationship between essay score and English grades might change based on type of English course examined, for example, a composition versus a literature course. Furthermore, research is planned to explore those students who may have discrepant performance on the essay versus multiple-choice items on the SAT writing section, and determine the additional information garnered for admission or placement by having the essay on the test for these students. This would be similar to a study by Shaw, Mattern, and Patterson (2011) that found that after controlling for relevant student characteristics and prior academic performance, an SAT critical reading-writing discrepancy still had an effect on first-year GPA as well as on English course grades in college. Also, the role of student's first versus best language on direct versus indirect writing assessments may provide useful information in understanding our increasingly linguistically diverse college applicants' writing abilities. This information may help to develop and target interventions to improve writing performance.

Conclusion

This study found that the SAT essay score provides useful and unique information in understanding how students will perform in college generally, and in their English course work, information that is independent of the other academic variables which we held constant. In addition, the SAT essay appears to demonstrate somewhat smaller demographic subgroup differences in performance than on the SAT critical reading and math sections. This information, taken together, indicates that the essay may be a currently underutilized, but valuable tool in college admission and placement. With so many students struggling with college-level writing (The National Commission on Writing in America's Schools and Colleges, 2003), it would seem worthwhile to further explore this academic tool, particularly because students who take the SAT and apply to college are already providing this information to admission offices.

References

Bandura, A. (1993). Perceived self-efficacy in cognitive development and functioning. *Educational Psychologist*, *28*, 117–148.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: Freeman.

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

Breland, H.M., Bridgeman, B., & Fowles, M.E. (1999). *Writing assessment in admission to higher education: Review and framework* (College Board Research Report No. 99-3; GRE Board Research Report No. 96-12R). New York: The College Board.

Breland, H.M., Kubota, M.Y., & Bonner, M.W. (1999). *The performance assessment study in writing: Analysis of the SAT II: Writing Subject Test* (College Board Report No. 99-4). New York: The College Board.

Bridgeman, B., & Lewis, C. (1991). *Sex differences in the relationship of Advanced Placement essay and multiple-choice scores to grades in college courses* (ETS Research Report No. RR-91-48). Princeton, NJ: Educational Testing Service.

Bridgeman, B., Pollack, J., & Burton, N. (2004). *Understanding what SAT Reasoning Test™ scores add to high school grades: A straightforward approach* (College Board Research Rep. No. 2004-4). New York: The College Board.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

College Board. (n.d.). *How SAT essays are scored*. Retrieved from: http://professionals.collegeboard.com/testing/sat-reasoning/scores/essay.

College Board. (2007). *2007 College-Bound Seniors: Total Group Profile Report*. New York: The College Board.

College Board. (2008). *2008 College-Bound Seniors: Total Group Profile Report*. New York: The College Board.

Kobrin, J.L, Deng, H., & Shaw, E.J. (2007). Does quantity equal quality? The relationship between length of response and scores on the SAT essay. *Journal of Applied Testing Technology*, *8*(1), 1–15.

Kobrin, J.L., Deng, H., & Shaw, E.J. (2011). The association between SAT prompt characteristics, response features, and essay scores. *Assessing Writing*, *16*, 154–169.

Kobrin, J.L., Patterson, B.F., Shaw, E.J., Mattern, K.D., & Barbuti, S.M. (2008). *Validity of the SAT for predicting first-year college grade point average* (College Board Research Report No. 2008-5). New York: The College Board.

Kobrin, J.L., Sathy, V., & Shaw, E.J. (2007). *A historical view of subgroup performance differences on the SAT Reasoning Test* (College Board Research Report No. 2007-7). New York: The College Board.

Mattern, K., Kobrin, J., Patterson, B., Shaw, E., & Camara, W. (2009). Validity is in the eye ofthe beholder: Conveying SAT research findings to the general public. In R.W. Lissitz (Ed.), *The concept of validity: Revisions, new directions, and applications* (pp. 213–240). Charlotte, NC: Information Age.

Mattern, K.D., Patterson, B.F., Shaw, E.J., Kobrin, J.L., & Barbuti, S.M. (2008). *Differential validity and prediction of the SAT* (College Board Research Report No. 2008-4). New York: The College Board.

Matzen, R.N., & Hoyt, J.E. (2004). Basic writing placement with holistically scored essays: Research evidence. *Journal of Developmental Education*, *28*(1), 2–12, 34.

Messick, S. (1989). Validity. In R.L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–103). New York: Macmillan Publishing Company.

Noeth, R.J., & Kobrin, J.L. (2007). *Writing changes in the nation's K–12 education system* (College Board Research Note No. RN-34). New York: The College Board.

Patterson, B.F., & Mattern, K.D. (2010). *Validity of the SAT for predicting FYGPA: 2008 SAT validity sample* (College Board Statistical Report). New York: The College Board.

Patterson, B.F., Mattern, K.D., & Kobrin, J.L. (2009). *Validity of the SAT for predicting FYGPA: 2007 SAT validity sample* (College Board Statistical Report). New York: The College Board.

Pomplun, M., Wright, D., Oleka, N., & Sudlow, M. (1992). *An analysis of English composition test essay prompts for differential difficulty* (College Board Research Report No. 92-4). New York: The College Board.

Shaw, E.J., Mattern, K.D., & Patterson, B.F. (2011). Discrepant SAT critical reading and writing scores: Implications for college performance. *Educational Assessment*, *16*, 145–163.

Shaw, E.J., & Patterson, B.F. (2010). What should students be ready for in college? *A look at first-year course work in four-year postsecondary institutions in the U.S.* (College Board Research Report No. 2010-1). New York: The College Board.

The National Commission on Writing in America's Schools and Colleges. (2003). *The neglected "R": The need for a writing revolution*. New York: The College Board.

Willingham, W.W., & Cole, N.S. (1997). *Gender and fair assessment*. Mahwah, NJ: Lawrence Erlbaum Associates.

Table 1.

The Relationship Between SAT Essay Score and Demographic Characteristics

		n	*M*	*SD*	*d*
Gender	Female	66,344	7.88	1.41	0.04
	Male	54,553	7.73	1.57	-0.06
Race/Ethnicity	American Indian	623	7.64	1.46	-0.12
	Asian	10,393	8.13	1.56	0.21
	Black/African American	8,614	7.15	1.44	-0.45
	Hispanic	9,637	7.43	1.50	-0.26
	Other	3,427	7.94	1.50	0.08
	White	82,850	7.87	1.45	0.04
	No Response	5,353	8.02	1.56	0.13
Highest Parental Education	No High School Diploma	2,602	7.16	1.48	-0.45
	High School Diploma	25,957	7.36	1.44	-0.31
	Associate Degree	7,965	7.45	1.39	-0.25
	Bachelor's Degree	39,758	7.85	1.44	0.02
	Graduate Degree	38,530	8.21	1.47	0.26
	No Response	6,085	7.80	1.54	-0.01
Best Language Spoken	English	111,520	7.84	1.48	0.01
	English and Another	6,488	7.63	1.56	-0.13
	Another	1,244	7.06	1.63	-0.51
	No Response	1,645	7.70	1.64	-0.08
First Language Spoken	English	98,948	7.82	1.47	0.00
	English and Another	13,568	7.86	1.56	0.03
	Another	6,812	7.71	1.60	-0.07
	No Response	1,569	7.77	1.63	-0.03
Total		120,897	7.82	1.49	

Note: For all subgroups, standardized differences (*d*) are calculated as (Subgroup Mean minus Total Mean)/Total Standard Deviation.

Table 2.

The Relationship Between SAT Essay Score and Academic Performance

		n	*M*	*SD*	*d*
SAT (Critical Reading + Math + Writing Multiple Choice)	5 (2110–2400)	5,062	9.56	1.34	1.17
	4 (1810–2100)	30,746	8.65	1.29	0.56
	3 (1510–1800)	51,536	7.77	1.25	-0.03
	2 (1210–1500)	29,222	6.97	1.27	-0.57
	1 (600–1200)	4,331	6.04	1.42	-1.19
HSGPA	Above A-	72,355	8.06	1.45	0.16
	A- to B+	23,137	7.66	1.44	-0.11
	B+ to B-	21,976	7.33	1.45	-0.33
	Below B-	3,429	6.82	1.49	-0.67
First-Year GPA	Above B+	42,473	8.28	1.43	0.31
	B+ to C+	59,183	7.71	1.43	-0.07
	C+ to D+	16,144	7.18	1.45	-0.43
	Below D+	3,097	6.79	1.48	-0.69
First-Year Eng GPA	Above B+	56,600	8.10	1.43	0.19
	B+ to C+	46,272	7.70	1.45	-0.08
	C+ to D+	12,989	7.29	1.53	-0.36
	Below D+	5,036	6.97	1.52	-0.57
Retention to Second Year	Yes	107,950	7.87	1.48	0.03
	No	12,947	7.37	1.50	-0.30
Total		120,897	7.82	1.49	

Note: For all subgroups, standardized differences (*d*) are calculated as (Subgroup Mean minus Total Mean)/Total Standard Deviation.

Table 3.

Mean FYGPA by SAT Essay Score, Grouped by Student Achievement Levels Based on HSGPA and SAT Scores (CR + M + W Multiple Choice)

	Below Average Achievers				Average Achievers				Moderately High Achievers				High Achievers			
SAT Essay Score	***n***	***Mean***	***SD***	***d***	***n***	***Mean***	***SD***	***d***	***n***	***Mean***	***SD***	***d***	***n***	***Mean***	***SD***	***d***
0	3	nr	nr	nr	15	2.26	0.95	-0.32	9	nr	nr	nr	7	nr	nr	nr
2	15	1.81	0.91	-0.28	17	1.79	0.90	-0.99	9	nr	nr	nr	4	nr	nr	nr
3	17	1.89	0.76	-0.17	48	2.37	0.60	-0.18	16	2.39	0.93	-0.79	3	nr	nr	nr
4	77	2.07	0.67	0.05	271	2.23	0.79	-0.37	99	2.52	0.74	-0.57	62	2.94	0.86	-0.91
5	99	1.97	0.78	-0.07	542	2.34	0.74	-0.21	252	2.66	0.69	-0.35	187	3.21	0.58	-0.36
6	238	2.00	0.80	-0.04	2,323	2.41	0.74	-0.12	1,390	2.74	0.64	-0.22	1,303	3.21	0.59	-0.35
7	161	2.07	0.76	0.06	2,259	2.50	0.68	0.01	2,008	2.83	0.65	-0.08	2,611	3.30	0.55	-0.18
8	115	2.08	0.77	0.06	2,895	2.56	0.69	0.10	4,267	2.91	0.60	0.05	9,713	3.36	0.49	-0.05
9	13	nr	nr	nr	637	2.65	0.65	0.22	1,718	2.97	0.58	0.15	7,198	3.41	0.46	0.05
10	2	nr	nr	nr	177	2.69	0.60	0.28	660	2.99	0.56	0.19	5,009	3.44	0.44	0.12
11	0	nr	nr	nr	25	2.75	0.54	0.36	142	2.96	0.56	0.13	2,017	3.48	0.45	0.20
12	0	nr	nr	nr	4	nr	nr	nr	43	3.03	0.50	0.24	740	3.50	0.40	0.25
Total	740	2.03	0.77		9,213	2.49	0.71		10,613	2.88	0.62		28,854	3.38	0.49	

Note: Groups with fewer than 15 students are not reported. High Achievers have an HSGPA ≥ A- and SAT (CR + M + W multiple choice) > 1800. Moderately High Achievers have an A- > HSGPA ≥ B+ and 1800 ≥ SAT (CR + M + W multiple choice) > 1500. Average Achievers have a B+ > HSGPA ≥ B- and 1500 ≥ SAT (CR + M + W multiple choice) > 1200. Below Average Achievers have a B- ≥ HSGPA and 1200 ≥ SAT (CR + M + W multiple choice) > 600. For all subgroups, standardized differences (*d*) are calculated as (Subgroup Mean minus Total Mean)/Total Standard Deviation.

Table 4.

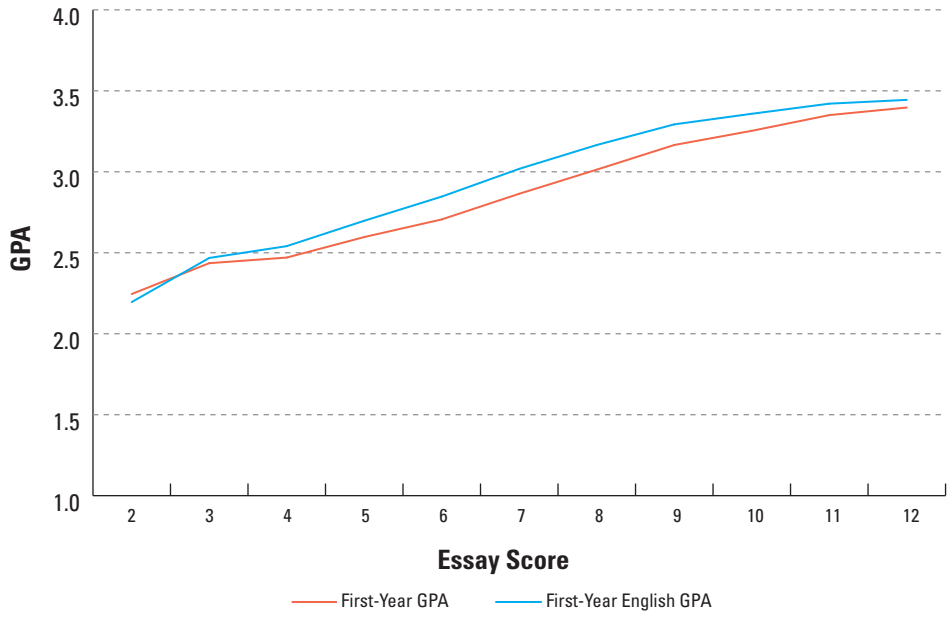
Mean FY EngGPA by SAT Essay Score, Grouped by Student Achievement Levels Based on HSGPA and SAT Scores (CR + M + W Multiple Choice)

	Below Average Achievers				Average Achievers				Moderately High Achievers				High Achievers			
SAT Essay Score	***n***	***Mean***	***SD***	***d***	***n***	***Mean***	***SD***	***d***	***n***	***Mean***	***SD***	***d***	***n***	***Mean***	***SD***	***d***
0	3	nr	nr	nr	15	2.30	1.16	-0.40	9	nr	nr	nr	7	nr	nr	nr
2	15	1.77	1.29	-0.32	17	1.73	1.13	-1.01	9	nr	nr	nr	4	nr	nr	nr
3	17	1.95	0.91	-0.14	48	2.50	0.93	-0.19	16	2.46	1.38	-0.76	3	nr	nr	nr
4	77	2.14	0.84	0.05	271	2.31	1.08	-0.39	99	2.58	0.96	-0.61	62	2.99	0.96	-0.68
5	99	1.84	0.98	-0.25	542	2.47	1.02	-0.22	252	2.76	1.00	-0.38	187	3.31	0.76	-0.19
6	238	2.07	1.05	-0.02	2,323	2.58	0.98	-0.11	1,390	2.92	0.87	-0.18	1,303	3.32	0.75	-0.17
7	161	2.21	1.10	0.11	2,259	2.71	0.90	0.03	2,008	2.98	0.83	-0.10	2,611	3.36	0.72	-0.12
8	115	2.22	1.01	0.12	2,895	2.76	0.91	0.09	4,267	3.11	0.75	0.06	9,713	3.42	0.64	-0.02
9	13	nr	nr	nr	637	2.87	0.83	0.20	1,718	3.18	0.72	0.15	7,198	3.45	0.61	0.04
10	2	nr	nr	nr	177	2.96	0.75	0.30	660	3.19	0.72	0.16	5,009	3.48	0.60	0.07
11	0	nr	nr	nr	25	3.05	0.92	0.40	142	3.21	0.70	0.19	2,017	3.50	0.61	0.10
12	0	nr	nr	nr	4	nr	nr	nr	43	3.19	0.74	0.16	740	3.51	0.60	0.13
Total	740	2.09	1.03		9,213	2.68	0.94		10,613	3.06	0.79		28,854	3.43	0.64	

Note: Groups with fewer than 15 students are not reported. First-year English GPA is the average grade of all English courses taken by a student in his or her first year of college. High Achievers have an HSGPA ≥ A- and SAT (CR + M + W multiple choice) > 1800. Moderately High Achievers have an A- > HSGPA ≥ B+ and 1800 ≥ SAT (CR + M + W multiple choice) > 1500. Average Achievers have a B+ > HSGPA ≥ B- and 1500 ≥ SAT (CR + M + W multiple choice) > 1200. Below Average Achievers have a B- ≥ HSGPA and 1200 ≥ SAT (CR + M + W multiple choice) > 600. For all subgroups, standardized differences (*d*) are calculated as (Subgroup Mean minus Total Mean)/Total Standard Deviation.

Figure 1.

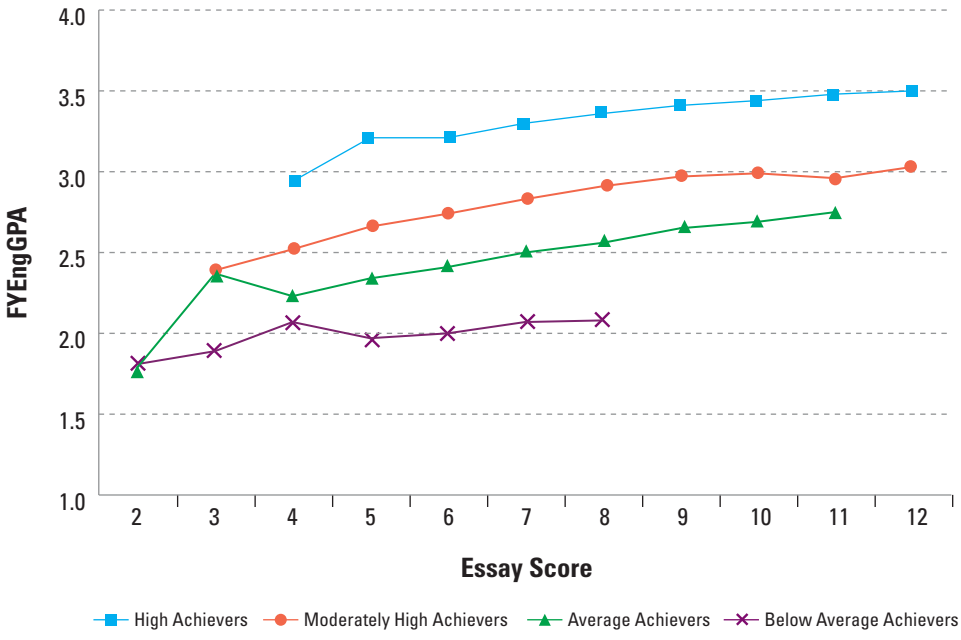
Mean FYGPA and FY EngGPA by SAT essay score.



Note: Essay scores of 0 are not plotted because a 0 score represents not following directions as opposed to poor performance.

Figure 2.

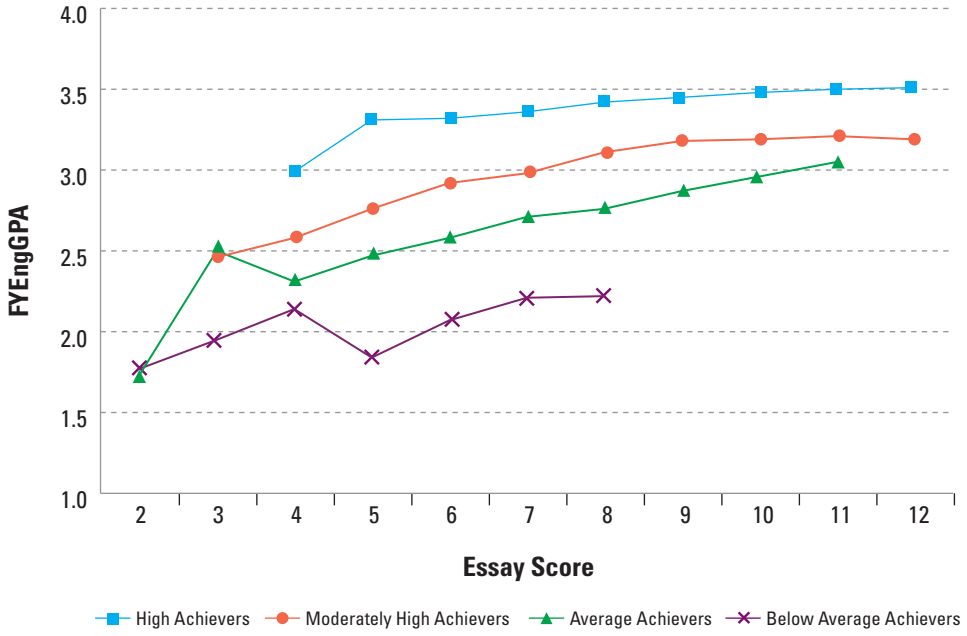
Mean FYGPA by SAT essay score for students at different achievement levels.



Note: High Achievers have an HSGPA ≥ A- and SAT (CR + M + W multiple choice) > 1800. Moderately High Achievers an have an A- > HSGPA ≥ B+ and 1800 ≥ SAT (CR + M + W multiple choice) > 1500. Average Achievers have a B+ > HSGPA ≥ B- and 1500 ≥ SAT (CR + M + W multiple choice) > 1200. Below Average Achievers have a B- ≥ HSGPA and 1200 ≥ SAT (CR + M + W multiple choice) > 600.

Figure 3.

Mean FY EngGPA by SAT essay score for students at different achievement levels.



Note: High Achievers have an HSGPA ≥ A- and SAT (CR + M + W multiple choice) > 1800. Moderately High Achievers have an A- > HSGPA ≥ B+ and 1800 ≥ SAT (CR + M + W multiple choice) > 1500. Average Achievers have a B+ > HSGPA ≥ B- and 1500 ≥ SAT (CR + M + W multiple choice) > 1200. Below Average Achievers have a B- ≥ HSGPA and 1200 ≥ SAT (CR + M + W multiple choice) > 600.

The Research department actively supports the College Board's mission by:

- Providing data-based solutions to important educational problems and questions
- Applying scientific procedures and research to inform our work
- Designing and evaluating improvements to current assessments and developing new assessments as well as educational tools to ensure the highest technical standards
- Analyzing and resolving critical issues for all programs, including AP®, SAT®, PSAT/NMSQT®
- Publishing findings and presenting our work at key scientific and education conferences
- Generating new knowledge and forward-thinking ideas with a highly trained and credentialed staff

Our work focuses on the following areas

Admission	Measurement
Alignment	Research
Evaluation	Trends
Fairness	Validity

Follow us online: research.collegeboard.org

CollegeBoard

CollegeBoard

13b-7395