Estimating the value of SAT Writing: Admissions and Student Success

Paper presented to the California Association of Institutional Research
Sacramento, CA

Bryce Mason, Ph.D.
Institutional Research
Loyola Marymount University
19 November 2009

The College Board revamped the Scholastic Aptitude Test (SAT) in 2005 to include a writing component, making it more of an achievement test than ever before, but institutions of higher education largely have been slow to do much with it citing the need for more analysis on its ability to predict student performance (Chute, 2007; Honawar, 2005). Through fall 2009, most institutions around the country apparently have gone three recruitment cycles without incorporating the new score into the admissions ethos. Nevertheless, three cohorts with writing scores have now gone through their first year of college, and very soon the first cohort with scores will be starting to graduate. Thus, researchers should be beginning to come to a consensus on the predictive value of this new test with respect to student success.

Dozens of studies have been conducted on the ability of the old SAT to predict achievement with the overwhelming majority of results confirming that high school grades (HSGPA) and SAT scores are statistically significant predictors and together account for a substantial proportion of the variance in first-year GPA (e.g., Geiser and Studley, 2001; Camara and Echternacht, 2000). In general, HSGPA has been found to have the largest standardized beta weight, with SAT math and verbal scores adding in a modest additional amount of information. Recently, however, at least two scholars found that SAT scores also measure socio-economic factors (SES), suggesting that the SAT's predictive validity has been overestimated over the preceding decades (Rothstein, 2004; Geiser and Studley, 2001). Nevertheless, SAT scores retained a smaller but still significant relationship to first-year GPA, even after controlling for SES and other factors. Thus, the old verbal and mathematics batteries appeared to have provided at least some value to admissions offices during their tenure.

Certainly, the changes to the SAT have been numerous in this latest iteration. First, the verbal test (now named critical reasoning) no longer contains analogies, instead replaced with shorter reading passages and multiple-choice questions to ascertain students' comprehension. Second, the mathematics section added third-year (algebra II) content in exchange for the previously used quantitative comparisons section. Finally, the test added a one-hour writing section that focuses on usage, grammar, and diction. It contains 35 minutes of multiple-choice questions to identify errors or to improve sentences and paragraphs, and a 25-minute essay question scored by trained college and high school teachers. The test now requires a whopping three hours and forty-five minutes, a 25% increase over the previous version. With the recent changes to the

SAT, institutions of higher education must reestablish the historical relationships between scores and student success in order to assess whether the measures should carry the same or any weight in admissions decisions.

Researchers need not go into this endeavor blind, however, as the new writing component has its roots in the optional SAT II subject-matter writing test. In its previous incarnation, it had a 20-minute essay and a 40-minute multiple-choice section to demonstrate English excellence. The University of California system has data on thousands of students over many years that elected to submit this score. As noted by Geiser (2008) the 2002 choice by UC to favor results on subject-matter tests over aptitude tests was grounded in decades of institutional data showing that the former performed at least as well as the latter—without the SES bias. Moreover, UC found that the writing subject test in particular was one of the best predictors available, second only to HSGPA (Geiser and Studley, 2001). These findings have much face validity, since writing is a key skill required in most major areas of study.

A review of the referred literature found no studies on the additional predictive validity of the new SAT, but the UC, the College Board, and a handful of other campuses around the country have begun to consider its performance. College Board studies found that, as in the past, unadjusted correlations between first-year GPA and SAT sub-scores exist in the 0.2 to 0.3 range and provide information in addition to HSGPA (Korbin, Patterson, Shaw, Mattern, and Barbuti, 2008). UC recently wrestled with whether to drop its policy of favoring subject tests, given that their most-recent institutional research showed that the new SAT I performed well enough without their inclusion (Agronow and Rashid, 2007). This policy proposal was not without its detractors, however (Geiser, 2008). Finally, scholars from the University of Georgia found that the new writing component outperformed the other two sub-tests for predicting first-year GPA, English grades, and the number of hours enrolled and withdrawn (Cornwell, Mustard, Van Parys, 2008).

Thus, Loyola Marymount University's Institutional Research (IR) office aimed to determine whether the SAT Writing (SATW) score provided *additional value* above and beyond what was already known about the student's prior performance (i.e., what was already included in the campus' admissions index, namely SAT verbal (SATV), math (SATM), and HSGPA). Two years of incoming LMU freshmen and their first-year outcomes were studied in order to determine whether the campus could improve upon its success metrics merely by reordering the admissions queue using previously unused information.

Method

The Admissions Office and the associate deans of LMU's six colleges asked IR to explore three hypotheses. First, if it was used, would the SATW score have provided additional predictive power in terms of five student and institutional outcomes of interest: (1) first-year GPA, (2) English 110 grade (the first English course), (3) academic probation status (<2.0 GPA) after the first semester, (4) first-to-second year retention, and (5) the likelihood of an admitted student

matriculating? Information gained with respect to this question would inform whether to incorporate the SATW score into the admissions index. Second, if any, were relationships uniform or did they differ by college? The answer to this question would help the associate deans of the colleges decide whether they could push for differing admissions standards by discipline. Finally, the Admissions Office was curious to know whether the commonly used policy of summing the maximum of each submitted sub-score to calculate an applicant's SAT total score was the best use of information. In other words, were maximums of the individual scores submitted the best indicator of success? If another type of score was superior, such as the first score submitted or the average score, that would provide information on whether current practice was aligned with the outcomes trying to be maximized.

Data

Data from two sources were combined to create the sets used for this analysis. First, Admissions provided individual-level records of all applicants to LMU for enrollment in fall 2006 and fall 2007. Each record included a name, high school GPA, all of the various SAT score submissions (e.g., a student took the SAT twice), student identification numbers, and the applicant's final status (i.e., admitted or matriculated). The multiple SAT scores were merged with the admissions data to identify the maximum SAT sub-scores, the first submitted sub-scores, and the average sub-scores for every individual. Second, the Registrar's Office provided basic demographic data, first-year GPA, an indicator of first-year retention, and the fall and spring grade history for the new freshmen students. These registration data yielded English 110 grades, total credit hours taken, and an indicator for academic probation (i.e., a first-year GPA of less than 2.0).

Fall 2006 and fall 2007 applicants were pooled together in order to estimate relationships. This yielded a sample size of 2144 students, far less than the approximately 2500 students enrolled during those two years. Of these 2144, only 1821 took English 110 during their first year. The main reason for the reduction in the sample was that not all students had a comparable HSGPA, since a modest number of students come from international high schools. A second reason was that a significant number did not submit SAT scores (i.e., they used the ACT). No imputation of SAT scores was conducted.

Models

Ordinary least squares regression was used to test for differences in first-year GPA and English 110 grades related to the variables under study. For these pseudo-continuous outcomes, the following two sets of covariates were included.

1. Transcript-based HSGPA and maximum SATM / SATV sub-scores
2. The covariates in Model 1 plus the maximum SATW sub-score

That is, IR took the approach of building up the model to investigate how the addition of SATW changed the relative relationship between existing covariates and the outcomes. Because the

policy question at hand was whether the SATW would be a beneficial predictor of student success to be included into the admissions index, Model 1 was essentially the admissions index.

Logistic regression was used to examine the relationship of the same sets of covariates above with the other, dichotomous outcomes. The only exception was the admissions model, which looked strictly at SAT scores.

Finally, the models were run by college and by type of SAT scores in order to determine whether relationships were uniform across college and whether the policy of using the maximum SAT score is most aligned with maximizing the outcomes of interest.

Results

First-year GPA

Table 1 displays the estimated coefficients of Models 1 and 2 on first-year GPA. Results demonstrated that SAT writing was strictly better than SATV in predicting first-year GPA. In Model 1, SAT verbal had a statistically significant relationship with first-year GPA; however, as soon as SAT writing was added, the predictive power of SATV was eliminated, suggesting that the verbal score was merely a modest proxy for writing ability. In addition, the writing score had a larger relationship than the verbal score. Model 2 suggested that a 100-point increase in SATW equated to about 0.14 of a grade point at the end of a freshman's first year, whereas Model 1 estimated that 100 points on the verbal score was worth an increase of 0.08. Also expected, HSGPA was far and away the most powerful predictor of first-year GPA.

Table 1. First-year GPA

	Model 1		Model 2	
	Coefficient	*t*	Coefficient	*t*
HSGPA	0.854	19.33***	0.830	18.79***
SATV	0.079	3.72***	-0.005	-0.18
SATM	0.091	4.62***	0.065	3.18**
SATW			0.141	5.03***
N	2144		2144	
R^2	0.187		0.196	

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ – SAT scores divided by 100

English 110

As one might have expected, the writing score was positively correlated with English 110 grades. Model 2 estimated that about a 200-point change in SAT writing might be worth about a plus or a minus in the grade. The magnitude of this relationship is roughly comparable to that of first year GPA in Table 1. Perhaps most surprisingly in this analysis is that the verbal score never predicted English 110 outcomes—even before the writing score was introduced.

Again, high-school GPA is the strongest predictor of performance. The same level of change in English 110 grades (about a third of a grade point) is estimated to be equivalent to about half a high-school GPA point. In other words, in this outcome, 50 points on the SAT writing score is "worth" about 0.1 high-school GPA.

Table 2. English 110

	Model 1		Model 2	
	Coefficient	*t*	Coefficient	*t*
HSGPA	0.734	12.62***	0.709	11.66***
SATV	0.033	1.16	-0.051	-1.44
SATM	0.039	1.56	0.012	0.45
SATW			0.146	4.02***
N	1821		1821	
R^2	0.086		0.097	

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ – SAT scores divided by 100

First-year Retention

Apparently the admissions index was not very predictive of retention, since it only accounted for an estimated 1% of the variance. Average partial effects (APE, instead of odds ratios) were estimated for the logistic regression models and included in Table 3. Again, HSGPA had the strongest relationship, and the SAT scores possibly contributed a small amount of additional information, although this finding became murkier with the inclusion of SATW, which removed any predictive power of SATV.

Table 3. First-year Retention

	Model 1		Model 2	
	Coef. (APE)	*z*	Coef. (APE)	*z*
HSGPA	0.073	3.39**	0.074	3.41**
SATV	-0.024	-2.26*	-0.021	-1.55
SATM	0.024	2.49*	0.025	2.50**
SATW			-0.005	-0.37
N	2144		2144	
Pseudo R^2	0.013		0.013	

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ – SAT scores divided by 100

Academic Probation

As expected, higher HSGPA and SAT scores predict a lower probability of being on academic probation. Table 4 presents the estimates from models 1 and 2 for the academic probation outcome. The otherwise average person with a relative advantage of 100 SAT writing points might be expected to have a 4% less chance to be on academic probation at the end of the first semester.

Interestingly, a similar equivalency between the marginal effects of SAT writing and high school GPA appears just as in predicting first-year GPA and English 110 grades. Again, about 50 SAT writing points and 0.1 high-school GPA both yield a similar effect on the outcome of interest.

Table 4. Academic Probation

	Model 1		Model 2	
	Coef. (APE)	*z*	Coef. (APE)	*z*
HSGPA	-0.170	-8.12***	-0.162	-7.77***
SATV	-0.005	-0.48	0.019	1.53
SATM	-0.030	-3.29**	-0.022	-2.33*
SATW			-0.040	-3.01**
N	2144		2144	
Pseudo R^2	0.066		0.072	

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ – SAT scores divided by 100

Admission Acceptance

Admission acceptance is not often considered as an outcome of interest, but it is clearly related to many universities' goals. If relationships exist that predict matriculation, one could admit smaller populations that are more likely to enroll and have zero impact on expected enrollments. However, this would decrease the admissions ratio (applied vs. admitted) and increase the acceptance ratio (admitted vs. enrolled), both of which are related to selectivity measures used in college rankings.

Given the limited number of correlates in this data, the model only accounted for about five percent of the variation in students' decisions to enroll. However, Table 5 (where average partial effects were listed) showed that all of the SAT test scores were significant in this outcome. This was the only analysis where SATV was still significant when controlling for SATW, which makes sense since the scores are correlated and most competitor universities currently only use SATV in their offer decisions. A 100-point increase in all SAT scores (i.e. 300 points overall) was associated with a 17% less chance of accepting the admissions offer. However, because HSGPA was not included in this model (it was not available for all applicants, just the enrolled students because of a data glitch), this correlation cannot be compared directly to other SAT score correlations. If it were included, the relationships would almost certainly attenuate. In addition, this dataset did not have data on other factors that might affect the matriculation decision, such as financial aid or distance from the University. Nevertheless, the story here is obvious—those students with relatively higher SAT scores often choose to decline the offer, presumably because they have other options.

Table 5. Admission Acceptance

	Model 1		Model 2	
	Coef. (APE)	*z*	Coef. (APE)	*z*
SATM	-0.078	-10.65***	-0.065	-8.54***
SATV	-0.084	-10.96***	-0.045	-4.55***
SATW			-0.062	-6.17***
N	7925		7925	
Pseudo R^2	0.044		0.048	

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ – SAT scores divided by 100

College-level and Other Results

College differences. IR ran separate models for the colleges (stratified based on the student's entering college) to determine whether there were differences between colleges in academic success outcomes. The interested reader may enjoy seeing the tables in the appendix, but the key takeaways were as follows.

- HSGPA was the most powerful indicator for all colleges.
- The colleges of Liberal Arts, Business Administration, and Science and Engineering had the strongest relationship between SAT scores and academic success. The colleges of Film and Television and Communication and Fine Arts had non-significant results, although this may have been due to a statistical power issue because of the smaller number of students in those colleges.
- Liberal Arts had the strongest association with SATW, but surprisingly it mattered for Business Administration and Science and Engineering in terms of first-year GPA as well.

SAT score aggregation methods. IR hypothesized that there may be value in choosing different methods of using the available SAT score information. Current admissions policy is to use the maximum sub score over all tests submitted to LMU. However, one could argue that perhaps the first or average score a better indicator of ability. There were very slight differences in the predictive power of these models when using different methods of handling multiple test administrations. With few exceptions, the average score predicted outcomes best, followed by the maximum score, then the minimum score, and finally the first score. However, these differences were small. From the best to the worst, the difference was only seven-tenths of a percent of variation explained.

Conclusions

Overall, SATW appeared to be superior to SATV as a predictor of academic outcomes. While neither measure came close to the power of HSGPA, SATW did predict student success better than the SATV almost universally. The lone exception to this rule was in terms of student retention, where SATV was statistically significant in the existing index and SATW did not contribute any information. Of course, no variable in that model had much practical significance, since the total variance explained was around 1%. In all models, once SATW was added, SATV became insignificant. In terms of first-year GPA, the correlation for the SATV score shrank in magnitude to be very close to zero, implying that the predictive power of SATW was strictly better.

These findings are consistent with the other limited research on the SATW score. Cornwell, Mustard, and Van Parys (2008) also found that SATW helped to predict first-year GPA; moreover, they found that the SATV score became insignificant as SATW was introduced. Similarly, the UC committee on admissions (BOARS) found that, based on research by Agronow and Rashid (2007), the SATW does well enough to warrant the removal of SAT II subject tests as a requirement for admissions (with the possible exception of engineering).

In terms of choosing students to maximize the academic outcomes studied, a superior ordering of students could be obtained by replacing SATV with SATW in the LMU admissions index. This would have the effect of altering the order of the students on the admissions index continuum.

Practically speaking, a marginal number of students near the cut-off point with relatively higher SATW scores would be swapped in for students who had higher SATV scores.

Works Cited

Agronow, S. & Rashid, M. (2007, November). *Examining the Predictive Value of the SAT Subject Exams in the Prediction of First Year UC GPA*. Report submitted to the University of California Board of Admissions and Relations with Schools, Oakland, CA.

Camara, W.J. & Echternacht, G. (2000). *The SAT I and high school grades: Utility in predicting success in college*. New York: The College Board.

Chute, E. (2007, February 25). Colleges still unsure how to use new SAT. *Pittsburgh Post-Gazette,* pp. A1.

Cornwell, C.M., Mustard, D.B., & Van Parys, J. (2008). *How does the new SAT predict academic achievement in college?* Unpublished manuscript, University of Georgia, Athens, GA.

Korbin, 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*. New York: The College Board

Geiser, S. & Studley, R. (2001). *UC and the SAT: Predictive validity and differential impact of the SAT I and SAT II at the University of California*. Oakland, CA: University of California Office of the President.

Geiser, S. (2008). *Not so fast! A second opinion on a University of California proposal to endorse the new SAT.* UC Berkeley: Center for Studies in Higher Education. Retrieved from: http://escholarship.org/uc/item/0x09n63m.

Honawar, V. (2005). Colleges hesitate to embrace SAT writing test. *Education Week 24*(39). pp. 2.

Rothstein, J.M. (2004). College performance predictions and the SAT. *Journal of Econometrics 121*(2004). pp. 297-317

Appendix: College-level Models

Table 6: First-year GPA

	BA	CF	FT	LA	SE
HSGPA	0.915	0.826	0.933	0.714	1.024
	(10.64)***	(7.29)***	(5.97)***	(8.61)***	(11.42)***
SATV	-0.032	0.014	-0.071	-0.004	-0.074
	(-0.59)	(0.19)	(-0.76)	(-0.07)	(-1.37)
SATM	0.082	0.063	0.030	0.130	0.193
	(2.01)*	(1.17)	(0.44)	(3.23)**	(4.42)***
SATW	0.132	-0.026	0.131	0.153	0.152
	(2.49)**	(-0.35)	(1.36)	(2.97)**	(2.67)**
N	542	345	158	673	426
R^2	0.205	0.154	0.231	0.193	0.317

t statistics in parentheses
$^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

Table 7: English 110

	BA	CF	FT	LA	SE
HSGPA	0.663	0.795	1.209	0.520	0.961
	(6.54)***	(5.22)***	(5.13)***	(4.68)***	(7.36)***
SATV	0.014	-0.006	-0.092	-0.086	-0.085
	(0.22)	(-0.06)	(-0.64)	(-1.27)	(-1.12)
SATM	-0.03	-0.003	0.003	0.055	0.061
	(-0.66)	(-0.04)	(0.03)	(1.04)	(1.04)
SATW	0.126	0.017	0.057	0.221	0.133
	(2.04)*	(0.17)	(0.37)	(3.19)***	(1.69)
N	483	300	111	570	357
R^2	0.100	0.087	0.206	0.075	0.150

t statistics in parentheses
$^{*}p < 0.05, ^{**}p < 0.01, ^{***}p < 0.001$

Table 8: Freshman-to-Sophomore Retention

	BA	CF	FT	LA	SE
HSGPA	0.057	0.071	0.005	0.123	0.001
	(1.34)	(1.14)	(0.07)	(3.01)**	(0.02)
SATV	-0.045	-0.061	<0.001	0.032	-0.041
	(-1.75)	(-1.59)	(0.01)	(1.26)	(-1.65)
SATM	0.013	0.011	-0.018	0.012	0.042
	(0.63)	(0.39)	(-0.58)	(0.62)	(2.16)*
SATW	-0.020	0.003	0.051	-0.019	0.017
	(-0.77)	(0.07)	(1.12)	(-0.75)	(0.66)
N	542	345	158	673	426
pseudo R^2	0.026	0.016	0.025	0.025	0.028

z statistics in parentheses
$^{*}p < 0.05, ^{**}p < 0.01, ^{***}p < 0.001$
Average partial effects reported as opposed to odds-ratio

Table 10: Academic Probation

	BA	CF	FT	LA	SE
HSGPA	-0.134	-0.160	-0.123	-0.129	-0.214
	(-3.74)***	(-3.60)***	(-2.33)*	(-3.89)***	(-4.42)***
SATV	0.002	0.033	0.010	0.013	0.041
	(0.08)	(1.23)	(0.31)	(0.64)	(1.46)
SATM	-0.001	-0.028	-0.008	-0.025	-0.056
	(-0.05)	(-1.46)	(-0.33)	(-1.56)	(-2.55)*
SATW	-0.013	0.010	-0.029	-0.052	-0.056
	(-0.60)	(0.36)	(-0.85)	(-2.42)**	(-1.92)
N	542	345	158	673	426
pseudo R^2	0.050	0.085	0.073	0.092	0.113

z statistics in parentheses
$^{*}p < 0.05, ^{**}p < 0.01, ^{***}p < 0.001$
Average partial effects reported as opposed to odds-ratio