Draft: October, 2009

The SAT as a Predictor of College Success: Evidence from a Selective University*

Kevin Rask
Department of Economics
Wake Forest University
Winston-Salem, NC 27109
E-mail: Raskkn@wfu.edu

Jill Tiefenthaler
Office of the Provost
Department of Economics
Wake Forest University
Winston-Salem, NC 27109
E-mail: jmt@wfu.edu

Abstract

The SAT, long considered an objective measure for comparing across college applicants, has come under increased scrutiny in recent years. The scoring scandal of 2005-2006, the addition of the writing component in 2005, which significantly increased the testing time, and the recent statement by the National Association for College Admission Counseling that "it is the view of the Commission that there may be more colleges and universities that could make appropriate admission decisions without requiring standardized admission tests such as the ACT and SAT" have kept the efficacy of the test in the news in recent years. The debate centers around how useful the SAT is in predicting college success and what role the test plays in limiting access to higher education. This paper uses improved econometric techniques and a rich data set from a highly selective college to build on the literature that examines the usefulness of the SAT in predicting college performance. We investigate the potential bias in using the reported college GPA as a measure of student success, ignored in previous work, by converting the raw GPA to a z-score, which uses a person's relative rank in a course as the indicator of performance and, therefore, severs the link between the course choice and the nominal grade. In addition, we are able to look at the predictive power of the SAT across different groups – aided vs. nonaided students, white students vs. students of color, and men vs. women. The results indicate that the SAT increases the predictor power in a full model, but only marginally at 3 percentage points (increasing the r-squared from 28% to 31%). In addition, we find that the SAT has more predictive power for certain groups – men, minority students, and aided students – and that it takes large differences in SAT scores (more than 100 points) to generate meaningful differences in the college GPA.

* We are grateful to Allin Cottrell and acknowledge the comments of seminar participants at Wake Forest University.

1 Introduction

Confidence in the SAT has declined in recent years. From the increase in testing time resulting from the addition of a writing component in 2005 to the scoring scandal in 2006 to the recent increase in the number of colleges that do not require the SAT for college admissions, the test which was once widely accepted as an objective measure for comparing across college applicants is increasingly under fire. The chief complaint against the SAT is that it is not the best predictor of college success but is highly correlated with parental education and income and, therefore, contributes to perpetuating the inequality in higher education. This criticism comes at a time when highly selective colleges are being called on to address the fact that the vast majority of their students continue to come from families of high socioeconomic status.

The debate on the SAT continues in the higher education community and selective institutions are also divided on the weight it should be given in the admissions decision. While some schools are dropping the SAT because of its potential bias, others are giving it increased weight in admissions decisions. For example, Drew University President, Robert Weisbuch wrote that his institution dropped the SAT in 2006 to improve the applicant pool and encourage high-quality students of color to apply and also because it showed little correlation with student performance[1]. Many college presidents at top liberal arts institutions apparently agree. Bates, Bowdoin, Hamilton, Holy Cross, Smith, and Middlebury have all made the SAT optional in recent years. In 2008, Wake Forest University became the first top-ranked university to go test-optional, followed by New York University in 2009. However, most selective colleges continue

[1] "A President's First Year: Pacing Myself," Robert A. Weisbuch, *The Chronicle of Higher Education*, 52, 12, C3.

to require the SAT and some are even giving it more weight in the admissions decision. For example, a few years ago Towson University announced an experimental program to enroll more young men that admits applicants who score higher on the SAT but have lower high school grades than their regular admits[2].

The growing debate on the SAT makes it clear that research is needed to examine the efficacy of the test. Is the SAT a good measure of student ability? Is it equally predictive across race, gender, and socioeconomic status? Are there other variables available in the admissions file that can be used instead? Perhaps variables that are less correlated with socioeconomic status? This paper uses improved econometric techniques and a rich data set from a highly selective liberal arts college to build on the literature that examines the usefulness of the SAT in predicting college performance. We investigate the potential bias in using the reported college GPA as a measure of student success, ignored in previous work, by converting the raw GPA to a z-score, which uses a person's relative rank in a course as the indicator of performance and, therefore, severs the link between the course choice and the nominal grade. In addition, we are able to look at the predictive power of the SAT across different groups – aided vs. nonaided students, white students vs. students of color, and men vs. women.

The paper is outlined as follows. Section 2 presents some background on the SAT and summarizes the relevant literature. The data, discussed further in Section 3, are a 17 year panel following students through their college careers from a highly selective small university. The descriptive results are presented in Section 4. In Section 5 the results of estimating two models to explain college performance are discussed. The first is a selection model that first predicts the probability of graduation. Conditional upon graduation we then estimate a GPA equation based

[2] "Towson U. Give Men With Low Grades a Chance at College," *The Chronicle of Higher Education,* 53, 21, A38.

upon ones graduating GPA. The second model utilizes the full sample to predict first-year GPA. The conclusions and suggestions for further work are outlined in Section 6.

2 Background & Literature Review

The Scholastic Aptitude Test (SAT) was developed by Carl Brigham, a Princeton psychologist, in the 1920s. It was initially commissioned by a Harvard Dean, Henry Chauncey, with a goal of identifying the brightest students from across the country. The test quickly became popular with many Ivy League admissions offices in their efforts to diversify their student bodies beyond the Northeastern elite. It has been administered by the Education Testing Service (ETS) since the 1940s. The SAT was quickly adopted by colleges for admissions decisions and nearly 400 schools required the test for admittance by the 1950s. The major attraction of the test was the notion that, unlike high school grades, it could be used to compare applicants from across the country, from different types of schools and backgrounds.

Although the SAT has enjoyed dominance in college admissions for decades, the test has also been the subject of much criticism and this criticism has gained steam in recent years. Much of the debate focuses, appropriately, on the value of the SAT in predicting college success, typically measured as first-year or graduating GPA. As far back as 1984, Crouse and Brams argued '…that the predictive usefulness of the SAT for college admissions policies cannot be defended very well because it is very small. The marginal improvement in the admission policies of the College Board's 2,500 participating colleges does not seem to justify the huge costs of testing.' There is significant evidence on the correlation between the SAT and college GPA in the psychology literature, well reviewed in Berry and Sackett (2009). There is also an economics literature that examines SAT correlations with economics course grades and cumulative grades (Cohn et.al., 2003; Grant, 2007; Elzinga and Melaugh; 2009). Geiser et. al (2002a, 2002b, 2006, 2007, 2008) has used the University of California system data to examine the validity of SATs, high school grades, and achievements for predicting academic performance and finds that high school grades and achievement tests perform better than the SAT in their system. Rothstein (2004) is one of the few studies that provide evidence about the relative value of the SAT in predicting GPA holding other important factors, such as high school GPA, constant.

Another set of studies on the importance of the SAT in making good admissions decisions result from research of the test-optional experience. In 1984, for example, Bates College made the SAT optional, and now about a third of each class enters without submitting an SAT score. In a 20-year study of their policy and its results, Hiss and Neupane (2004) found that the difference in the performance of the SAT submitters and non-submitters is not significant (GPA average of 3.06 for non-submitters and 3.11 for submitters). The difference in Bates' graduation rates between submitters and non-submitters is one-tenth of one percent (0.1%). Robinson and Monks (2004) also found small differences in GPAs among Mt. Holyoke first years after they went test-optional in admissions.

One of the major attractions of the SAT when it was adopted by Harvard in the 1930s was the notion that it was comparable across students from a vast number of high schools across the country. However, as the importance of the test has grown over the years, its objectivity appears to have suffered as the industry around the SAT has grown as well. There is growing evidence of strategic behavior in test taking and also some evidence that the behavior pays off by increasing scores (Vigdor and Clotfelter, 2003; Robinson and Monks, 2004; Rothstein, 2009; Bound, et. al. 2009). Bound et al. (2009) find that as the number of high school graduates applying to college has risen (in the face of very small increases in admissions slots, especially at the most highly-ranked schools), there is evidence of increased test taking, increased test re-taking, and finally, in those areas where college admissions have become most competitive, a significant increase in the number getting to take the test under 'nonstandard' conditions. They also chronicle increases in those engaging in test preparation, both through private classes, private tutoring, and test preparation of any form.

In this paper, we look at the predictive power of the SAT using panel data from a highly selective university. We also examine the predictive power by group – aided vs. full-pay, men vs. women, and white students vs. students of color – to determine if it adequately predicts for some groups but not others. Much of the literature uses selected samples and self-reported data. In addition, all previous work relies on the college GPA as the measure of success. However, this variable may be a biased indicator as student major choice dramatically impacts their GPA. If weaker students are drawn to less demanding courses and/or majors or less ambitious students are drawn to courses and/or majors with relative grade inflation, measures of student ability will lose some of their predictive power. In this paper, we examine the predictability of variables available in the admissions file on college success of students enrolled at a selective university over 16 years. We investigate the potential bias in using the reported college GPA as a measure of student success, ignored in previous work, by converting the raw GPA to a z-score, which uses a person’s relative rank in a course as the indicator of performance and, therefore, severs the link between the course choice and the nominal grade.

In addition, we are able to look at the predictive power of the SAT across different groups. If SAT prep courses and video games do give wealthy students an advantage, it is likely the SAT has strong predictive power across wealthy students and across aided students but not across the entire group. Also, some argue that scores are biased against women. If that is the case SATs may have more predictive power for men than women.

3 Empirical Results

The data are derived from the cohorts entering fall 1992 through fall 2008. We initially looked at two main samples, the full sample used to predict first-year GPAs, and a somewhat smaller sample consisting only of graduates that are used for models of graduating GPA. Because the results are largely similar, our modeling results in section 3.2 are based solely on the graduating sample.

3.1 Descriptive Results

The samples consist of all new first-year students who entered beginning with the fall 1992 cohort. To be included in the first-year sample they had to complete their first year, and they had to graduate by the spring of 2008 to be included in the graduated sample. Both samples are similar in their characteristics, except for the smaller number and higher GPAs of the graduating sample.

Table 1: Descriptive Statistics of Graduated Sample and First-Year Sample

Group	Mean	75th Percentile	25th Percentile
Obs: Grads/FYs: 8,529 / 11,724			
GPA	3.16 / 2.98	3.44 / 3.37	2.91 / 2.67
SAT	1292 / 1299	1370 / 1380	1230 / 1230
Female	51.7% / 50.9%		
Minority	8.5% / 9.8%		
Financial Aid	43.7% / 40.6%		

Of primary interest is the relationship between ones SAT score and ones GPA. Figures 1 and 2 illustrate simple scatter diagrams for these. From both figures it is clear that there is a relationship between the two. These are the kinds of scatter diagrams that lead to correlation coefficients like those reported in both the psychology and the economics literature.

Figure 1

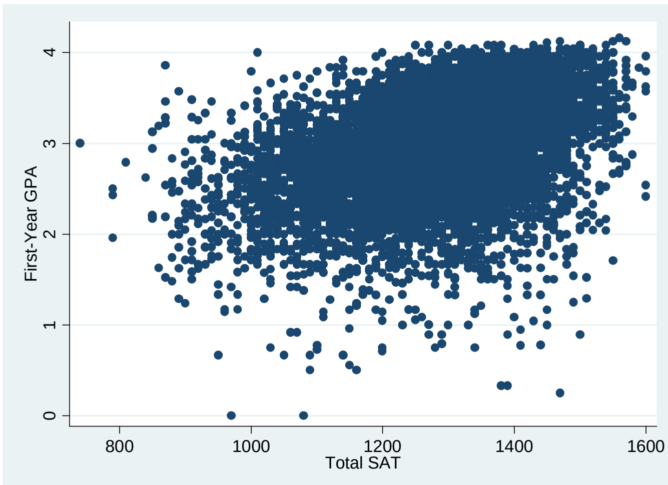
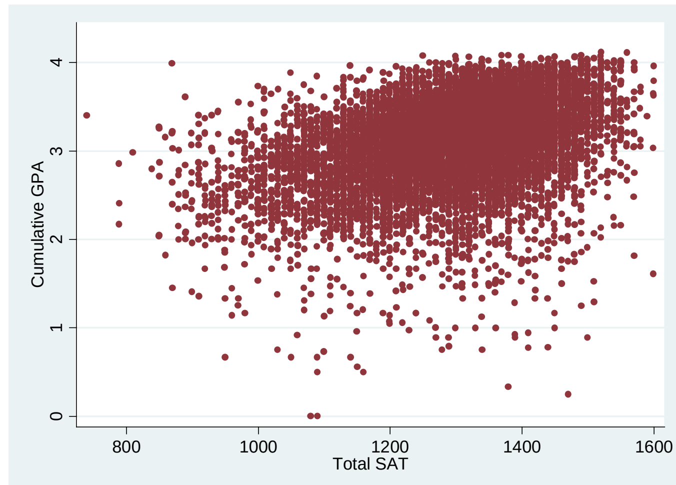


Figure 2



However, it is also clear from these illustrations that there are two other important characteristics of the relationship. First, it is very imprecise. There is significant scatter, high variance, whether you are looking at graduating GPA or first-year GPA. The second important feature of these graphs is that the slope is very flat. They are upward sloping, signifying the positive relationship between the two. However, the actual difference between the average GPAs of people with quite different SAT scores is not very large. Figures 3 and 4 take a vertical slice out of Figure 2 at two points, those with 1200 and those with 1400 SATs to further illustrate these issues. It is clear that the average GPA for those with a 1200 SAT is below that for those with a 1400. However, the variability is also clear. There are a significant portion of 1200 SAT people with GPAs between 2.0 and 3.0, just as there are for the population with 1400 SATs. These issues of variability and slope are explored more formally below.

Figure 3

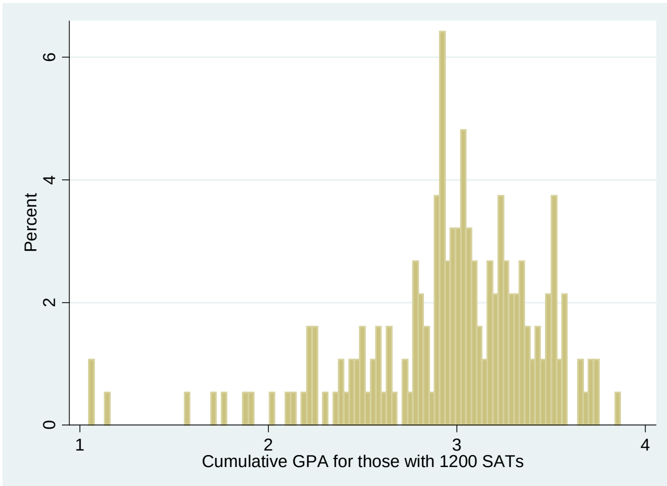
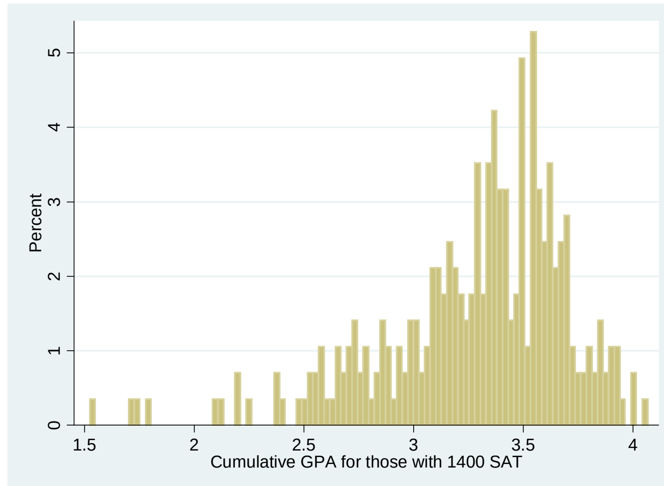


Figure 4



3.2 Empirical Model and Results

We evaluate the marginal explanatory power of the SAT through a series of GPA equations that utilize readily available admissions information. Because SATs, high school grades, and APs all have some overlap in what they measure, our strategy is to use multivariate regression analysis to illustrate the independent explanatory power of each. We report the R^2 from these equations as the measure of the proportion of variance across GPAs explained by the factors in each model. The models range from simple regressions with high school grades, or SATs, for example, to multiple regressions that add gender, race/ethnicity, and region of the country. Equations 1-4 illustrate some of the models estimated.

(1) $GPA = A + \beta_1 HS - GPA$

(2) $GPA = A + \beta_1 SAT$

(3) $GPA = A + \beta_1 APs$

(4) $GPA = A + \beta_1 HS - GPA + \beta_2 SAT + \beta_3 APs + \beta_4 Gender + \beta_5 Race + \beta_6 HomeRegion$

3.2.1 Explanatory Power of the SAT

Table 2 contains the R^2 from these models. For example, in column (1) the results from a simple model of graduating GPA vs. SAT are reported. In the full sample, the SAT explains

15% of the variability across college GPAs. Similarly, HS grades explain 17% and AP credits 11%. The significant amount of overlap in these measures of academic ability is evident in column 4. It contains the results from a multiple regression model that controls for all three factors. In total, those three factors explain 27% for the full sample. To assess the marginal benefit of the SAT, column 5 contains the results of model 4 after dropping the SAT from the equation. For the full sample the R^2 goes from 27% down to 22%, a loss of 5 percentage points of explanatory power. To further investigate the overlap we estimate a more complete model that includes race, gender, and region of the country. This 'full' model explains 31% of the variability in the full sample. Finally, again dropping the SAT from the full model gives the results in column 7, where we lose 3 percentage points in explanatory power by not knowing someone's SAT in the admissions process.

Table 2: R^2 from OLS Models of Grad-GPA & Academic and Socio-Demographic Characteristics

Group	(1) SAT	(2) HS-GPA	(3) APs	(4) All Acad	(5) (4)-SAT	(6) Full Model	(7) (6)-SAT
Full Sample: 8,362	15%	17%	11%	27%	22%	31%	28%
Men: 3,973	19%	14%	12%	27%	21%	28%	23%
Women: 4,389	16%	15%	12%	27%	22%	27%	24%
White: 7,637	12%	16%	11%	24%	22%	28%	25%
Minority: 1,013	15%	12%	8%	23%	17%	28%	20%
Full-Pay: 4,667	6%	12%	9%	19%	17%	24%	22%
Aided: 3,695	25%	21%	14%	35%	27%	38%	33%
<u>Aided:</u>							
Parent Income-Q1	23%	25%	12%	38%	30%	43%	38%
Parent Income -Q2	25%	21%	14%	35%	27%	38%	33%
Parent Income -Q3	21%	21%	13%	32%	26%	36%	31%
Parent Income -Q4	18%	21%	15%	31%	28%	35%	31%

Looking at the various subgroups shows significant heterogeneity among the influences. The biggest difference is among aided and full-pay students. For full-pays the SAT on its own explains 6% of the variability in GPAs. When more inclusive models are estimated, the marginal explanatory power of the SAT for full-pays drops to 2 percentage points. It holds a little more predictive power for men as opposed to women; however, the marginal explanatory power of the SAT is similar across gender. In terms of race, the SAT holds more marginal explanatory power for minorities when compared to whites. Dropping the SAT from a GPA equation for minorities leads to an 8 percentage point drop in R^2 for minorities, while the effect for whites is only 3 percentage points. The link between SATs and income is also evident in the lower panel of Table 2. While the marginal explanatory power is largely similar across the aided income quartiles, the total explanatory power of the GPA equations is larger for the lower income quartiles. The results in Table 2 illustrate quite vividly what Brigham, the man responsible for the development of the original SAT, came to realize not long after it began to be widely used. Lehmann (1999) reports that in an unpublished manuscript Brigham wrote that the standardized testing movement was based on "one of the most glorious fallacies in the history of science, namely that the tests measured native intelligence purely and simply without regard to training or schooling. The test scores very definitely are a composite including schooling, family background, familiarity with English and everything else."

3.2.2 GPA Predictions from the SAT

Explanatory power is one way to evaluate the usefulness of the SAT in the admissions process. Evaluating the actual differences in predicted GPAs across students gives one a sense of the magnitude of the impact of SAT differences on expected academic outcomes. From the multivariate regression models we can confidently assert that SATs are positively correlated with both first-year and graduating GPAs. An important policy question that comes from this is what is the expected difference in GPA when comparing applicants with different SATs? For example, two applicants look similar ‘on paper’, but one has an 1100 and another has a 1200 SAT score. What is our best guess at the likely difference in their collegiate GPAs? Table 3 contains the slope estimates from our models of GPA determination that include the SAT.

Table 3: $\widehat{\beta_{SAT}}$ from OLS Models of Graduating-GPA

Group	SAT	SAT,GPA,AP	Full Model
Full Sample:	.0013	.0008	.0008
Men:	.0015	.0010	.0009
Women:	.0012	.0008	.0007
White:	.0012	.0007	.0008
Minority:	.0011	.0008	.0008
Full-Pay:	.0009	.0004	.0005
Aided:	.0015	.0010	.0010

The first column contains simple regression results, and the changes from the first to the second and third column illustrate the bias inherent in univariate analysis. Because of the correlation between SATs and high school grades, models with SAT as the only explanatory variable suffer from omitted variables bias as some of the effect of high school grades is being attributed to the SAT. When the other main factors are accounted for (columns 2 and 3) the results are remarkably stable across specifications. Using the full sample results from the full model, a β estimate of .0008 implies that the expected difference in GPA between our hypothetical people with 1100 and 1200 SAT scores is .08. To put this in some context, estimates of the gender difference in GPA in these models, the average amount that a woman's GPA differs from a man's holding the other included factors constant, is around .17. That is more than double the effect when compared to a 100-point difference in SAT. Stated in another way, it takes about a 200 point difference in SAT scores to erase the gender deficit that men face in college GPAs.

3.2.3 Confidence Intervals for the Mean GPAs by SAT Group

The results from the previous section highlight the average difference in estimates of collegiate GPA based upon differences in SAT scores. In this section we use the error inherent in average GPAs to illustrate the confidence with which we can identify differences in applicants' academic outcome based upon their SAT scores. For example, earlier in Figures 3 & 4 we illustrated a broad range of collegiate academic outcomes from those with 1200 and those with 1400 SATs. Looking at the ranges, there is significant overlap in GPAs between these two groups of students. In this section we formalize this comparison by calculating 95% confidence intervals for GPAs from a set of SAT scores. We begin with all those with a score of 1000 and then increment it by 50 up to the final group, those with an SAT score of 1500. The first column in Table 3 contains the average graduating GPA for each group. It is clear that as the SAT increases, the average GPA also increases, from a 2.58 to a 3.39 for the range illustrated here. However, looking at the lower and upper limits of a 95% confidence interval around each mean, we can see that the mean GPA for people with a 1000 is indistinguishable from those with an 1150. The upper limit (2.74) for those with a 1000 is higher than the lower limit (2.71) for those with an 1150. That example is the broadest range where outcomes are statistically identical at the 95% level, but other groups are identical in the 50-100 point range.

Table 3: Mean GPA and 95% Confidence Intervals by SAT Group

SAT Group: (n)	Mean GPA	Lower Limit	Upper Limit
1000: (31)	***2.58***	2.43	2.74
1050: (62)	***2.71***	2.58	2.84
1100: (74)	***2.77***	2.65	2.88
1150: (97)	***2.81***	2.71	2.92
1200: (202)	***2.95***	2.88	3.02
1250: (289)	***3.11***	3.07	3.15
1300: (518)	***3.16***	3.12	3.19
1350: (424)	***3.19***	3.15	3.23
1400: (305)	***3.28***	3.23	3.33
1450: (172)	***3.28***	3.21	3.35
1500: (77)	***3.39***	3.28	3.50

The results in Table 3 suggest admissions offices shouldn't let modest differences in SAT scores unduly influence their decisions. Evaluations and groupings should be made between applicants with 100-200 point differences rather than between those with 10 or 20 point differences. One should consider groups with broader ranges of scores as applicants likely to have equivalent academic collegiate outcomes.

4 Conclusions

The SAT, long considered an objective measure for comparing across college applicants, has come under increased scrutiny in recent years. The debate centers around how useful the SAT is in predicting college success and what role the test plays in limiting access to higher education. This paper uses improved econometric techniques and a rich data set from a highly selective college to build on the literature that examines the usefulness of the SAT in predicting college performance. We investigate the potential bias in using the reported college GPA as a measure of student success, ignored in previous work, by converting the raw GPA to a z-score, which uses a person's relative rank in a course as the indicator of performance and, therefore, severs the link between the course choice and the nominal grade. In addition, we are able to look at the predictive power of the SAT across different groups – aided vs. nonaided students, white students vs. students of color, and men vs. women.

The results indicate that the SAT increases the predictor power of the full model, but only marginally at 3 percentage points (increasing the r-squared from 28% to 31%). In addition, we find that the SAT has more predictive power for certain groups – men (compared to women), minority students (compared to white students), and aided students (compared to students with no financial aid). We also find that it takes large differences in SAT scores (more than 100 points) to generate meaningful differences in the college GPA. For example, it takes about a 200 point difference in SAT scores to erase the underperformance of men in terms of college GPAs.

In summary, our results show that the SAT has only marginal predictive power in explaining college success, measured as the graduating z-scored GPA, across all students; adding only 3 percentage points to the total predictive power of the model once other student characteristics available in the admissions file are accounted for. In future work, these marginal gains in predictive power should be evaluated against the costs of the SAT in evaluating its usefulness for college admissions. Interestingly, the SAT has more predictive power in explaining college GPA within certain groups than it does across all students. For example, the SAT is much more valuable in predicting college success for students on financial aid than it is for students who don't receive need-based aid. It is likely that students from high income families benefit from test preparation that diminishes the value of the test in predicting college performance.

Our results also indicate that caution should be used when comparing small differences in SAT scores to make college admissions decisions. We find that it takes large changes in SAT scores to generate meaningful increases in the college GPA. For example, for two otherwise identical applicants with SAT scores of 1100 and 1200, the student with the 1200 is predicted to have a .08 higher graduating GPA. In these types of equations the gender difference in GPA is more than double this SAT difference, with women outperforming men by about .17. Therefore, colleges that choose to use the SAT for evaluating applicants for admissions should be very careful not to make these decisions on the basis of small differences in SAT scores.

VI. REFERENCES

Berry, Christopher and Paul Sackett (2009) Individual Differences in Course Choice Result in Underestimation of the Validity of College Admissions Systems, *Psychological Science*,

Bound, John, Brad Hershbein, and Bridget Terry Long (2009) Playing the Admissions Game: Student Reactions to Increasing College Competition, NBER working paper #15272, August.

Cohn, Elchanan, Sharon Cohn, Donald Balch, and James Bradley Jr (2003) Determinants of undergraduate GPAs: SAT scores, high-school GPA and high-school rank, *Economics of Education Review*, Vol. 23, pp. 577-586.

Crouse, James and Marvin Brams (1984) Benefits and Costs of the SAT for College Admissions Policies, *Atlantic Economic Journal*, Vol. 12, No. 2, p. 80.

Elzinga, Kenneth and Daniel Melaugh (2009) 35,000 Principles of Economics Students: Some Lessons Learned, *Southern Economic Journal*, Vol. 76, No. 1, pp. 32-46.

Geiser, S. with R. Studley (2002a) UC and the SAT: Predictive validity and differential impact of the SAT I and SAT II at the University of California. *Educational Assessment*, vol. 8, no. 1, pp. 1-26.

Geiser, S. and R. Studley (2002b) Research addendum: Additional findings on UC and the SAT. Oakland, CA: UC Office of the President: www.ucop.edu/news/sat/resaddendfinal.pdf

Geiser, S. and M. Santelices (2006) The role of advanced placement and honors courses in college admissions. Pp. 75 -114 in P. Gandara, G. Orfield and C. Horn, (eds.), *Expanding Opportunity in Higher Education: Leveraging Promise*. Albany, NY: SUNY Press.

Geiser, S. and M. Santelices. (2007). Validity of high-school grades in predicting student success beyond the freshman year: High-school record vs. standardized tests as indicators of four-year college outcomes. Center for Studies in Higher Education, UC Berkeley: cshe.berkeley.edu/publications/publications.php?id=265.

Geiser, S. (2008). Back to the basics: In defense of achievement (and achievement tests) in college admissions. Research and Occasional Paper Series, Center for Studies in Higher Education, UC Berkeley. cshe.berkeley.edu/publications/publications.php?id=313.

Grant, Darren (2007) Grades as information, *Economics of Education Review*, Vol. 26, pp. 201-214.

Lemann, N. (1999) *The big test: the secret history of the American meritocracy.* Array New York: Farrar, Straus and Giroux.

Robinson, Michael, and James Monks (2004) Making SAT scores optional in selective college admissions: A case study, *Economics of Education Review*, Vol. 24, pp. 393–405.

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

Vigdor, Jacob, and Charles Clotfelter (2003) Retaking the SAT, *Journal of Human Resources*, Vol. 38, No. 1, pp. 1–33.