Macalester College
DigitalCommons@Macalester College

Honors Projects | Economics Department

6-9-2009

Does Sorting make a difference in the SAT's prediction of Course Grades and Cumulative GPA at Macalester College?

Yuen Ting Liu
Macalester College, vivianytl@gmail.com

Follow this and additional works at: http://digitalcommons.macalester.edu/economics_honors_projects

 Part of the Econometrics Commons, and the Educational Assessment, Evaluation, and Research Commons

Recommended Citation
Liu, Yuen Ting, "Does Sorting make a difference in the SAT's prediction of Course Grades and Cumulative GPA at Macalester College?" (2009). *Honors Projects*. Paper 19.
http://digitalcommons.macalester.edu/economics_honors_projects/19

This Honors Project is brought to you for free and open access by the Economics Department at DigitalCommons@Macalester College. It has been accepted for inclusion in Honors Projects by an authorized administrator of DigitalCommons@Macalester College. For more information, please contact scholarpub@macalester.edu.

Does Sorting make a difference in the SAT's prediction of Course Grades and Cumulative GPA at Macalester College?

Vivian Yuen Ting Liu
Macalester College

Abstract

This paper estimates the predictive power of the SAT for college success. Most studies use GPA as a proxy for college success, but GPA is subject to differing grading standards across instructors and departments. Using individual course grades of Macalester College students from 1996 to 2005, this paper uses fixed effects to control for students sorting themselves into particular departments and courses. After controlling for sorting by major, a general trend emerges: the Math SAT is more predictive for science and math-related social science majors while the Verbal SAT is more predictive for language majors. After controlling for sorting by course, the predictive power of both the SAT and the Math SAT increases with statistical significance. The results suggest that the Verbal SAT does not carry the same predictive power as the Math SAT. Furthermore, the predictive ability of the high school GPA drops by at least 10 times. The result implies that the SAT is a predictive and reliable indicator for college grades even though the SAT does not predict each major's grades equally. As a result, the SAT should not be the only emphasis in admission process, but neither should it be ignored.

Advisor: Karine Moe
Economic Department
Honor Thesis

I. Introduction

The SAT Reasoning Test (SAT) was first designed in 1926 as an aptitude indicator in college admission (Lemann, 1999).[1] It was not meant to measure achievement, so that all students would have an equal chance to enter elite universities despite their ethnicity and income. More than a century later, the SAT became the most popular standardized admission test. In 2008, 1.5 million high school students took the SAT in preparation for college admission. On the other hand, the SAT has also become the most controversial admission test. Many researchers (Goldman 1976; Page 1985; Crouse, 1988; Fincher, 1974; Armstrong, 2003) and news reporters (Pope, 2008; Rimer, 2008; Rimer, 2008) have developed concerns about the effect, efficiency, and equity of using the SAT as a predictor of college success. These researchers have estimated the face validity, reliability, and internal consistency of the SAT in predicting college grade point average (GPA) and retention. ¶ ? This paper draws attention to the SAT's prediction for college grades. While most researchers choose GPA as a general indicator of college success, overall GPA masks differences in grading practices between professors and academic departments. Even through the SAT and grades may have a positive relationship within a course, this relationship is distorted between courses since different professors and departments interpret the GPA scale differently. This paper, therefore, looks at the predictive ability of the SAT, taking into account the sorting of students into classes and departments.

Formerly called the Scholastic Aptitude Test and Scholastic Assessment Test, the SAT Reasoning Test is the most used standardized admission test in the United States, especially on the coasts. The Verbal and Math sections of the SAT assess students' college readiness in the

[1] As of today, the official name of the SAT is the SAT Reasoning Testing. "SAT" does not stand for anything.

area of analysis, critical thinking, and problem solving.[2] The College Board advises that admission officials use the SAT in combination with high school grades to predict the degree of students' academic success. Although high school grades may be a better reflection of applicants' academic achievement, the various high school curriculums make them difficult to interpret. As a result, the SAT becomes the common yardstick for admission offices in assessing applicants.

According to the National Association for College Admission Counseling Admission Trends Survey (NACAC, 2008), more universities and colleges have emphasized the importance of admission test scores in recent years. Sixty percent of the responding colleges considered admission test scores an important factor in the admission decision in 2006, compared to only forty-six percentage in 1993. This percentage in 2006 is higher than the attribute "considerable importance" given to grades in all high school courses (51%), admission essay (28%), class rank (23%), teacher's recommendation (20%), and interview (10%). Nevertheless, 76% of colleges saw strength of high school curriculum as an "important" factor in the admission decision.

The SAT has achieved a nearly iconic status with powerful effects on public policy, social mobility, and even individual identity (Gould, 1996). Examining the implicit assumption of the SAT—predictive ability for college success—is therefore important. In September 2008, William R. Fitzsimmon, the dean of admissions at Harvard, spoke at the Seattle NACAC Conference. He challenged colleges and universities to examine their use of standardized admissions tests and to consider whether the benefits (genuine predictive power and convenience of using standardized admission test) outweigh the disadvantages (gender and ethnic inequality in college access). He further urged colleges to make the admissions tests optional and called on *U.S. News & World Report* to drop the SAT from the college ranking formula. An increasing

[2] The College Board added the Writing section in 2005.

number of colleges and universities have made the SAT optional for college entrance, for example, Bates College, Smith College, Northwestern University, Middlebury College, Trinity University, and 280 more institutions.[3] Critics pointed out that SAT-optional colleges adopt the policy in order to increase college ranking (Stecklow, 1995; Yablon, 2001). Since the SAT-optional policy encourages high test scorers to submit SAT scores and attracts a larger applicant pool, the policy inflates the colleges' average SAT scores and makes the colleges seem more selective.

Regardless of the test-optional debate, researchers and educators have generally agreed that the SAT provides an efficient tool for comparison and has strong correlation with freshman grades (Bridgeman, MaCamley-Jenkins, and Ervin, 2000; Weitzman, 1982; Zwick and Sklar, 2005). Nevertheless, they debate whether the strong correlation still holds after taking into account high school GPA and other students characteristics (age, gender, ethnicity, and family background). The scores gap across gender, income and ethnicity also raises a debate about the SAT's fairness. The debate regarding the SAT's predictive ability remains as researchers continue to refine their empirical research methodology.

Despite decades of research, scholars have largely overlooked the effect that sorting has on SAT validity studies. Professors have different grading standards and systems between and even within departments. Students also self-select into majors or classes according to the classes/ majors' levels of difficulty, in addition to their academic drive. If this sorting effect is present, it would affect the interpretation of the GPA. For example, if a high SAT scorer chooses challenging majors or classes, the SAT scores may overestimate his or her grade; on the other hand, the SAT may underestimate a low SAT scorer's grade, if he or she chooses less demanding majors/classes. As a result, the between-department or between-course variation could distort the

[3] For full SAT-optional list see: http://www.fairtest.org/files/OptionalPDFHardCopy.pdf

SAT's prediction. Using full transcript data, this paper is the third to use course level data in the SAT literature.[4] It is also the first to examine the SAT's predictability ability (for both Math and Verbal) for academic success (in terms of course grades and cumulative GPA), controlling for individual courses across different disciplines.

The paper unfolds as follows: Section II reviews related literature to compare their methodologies and results. Section III develops the model for this paper. Section IV summarizes the data. Section V describes the empirical approach and analyzes the result. This section also discusses the implication of the results and direction for future research. Section VI concludes the paper.

II. Literature Review

The SAT literature has existed for more than two decades. Researchers debate whether or not the SAT predicts college success equally well across demographic groups. Much of the literature has focused on how accurately and how differently the SAT predicts college academic success for students of various races, socioeconomic status, and high school records. These studies have not, however, considered the effect of sorting within universities on their results.

The literature examines the predictive ability of the SAT in different ways: bivariate correlation, multiple correlation, regression coefficients, incremental validity, and mean level difference by demographic groups. Virtually all College Board-sponsored research found that the SAT is a good predictor for freshman GPA (FGPA). Studying three freshman classes in 26 colleges, Bridgeman, Pollack, and Burton (2008) found that the coefficient of correlation between SAT scores and freshman grades is 0.55. The College Board's *Handbook for the SAT*

[4] Bridgeman and Wendler (1989) use course-level data to look at the Math SAT's prediction of grades in college mathematics courses. Bridgeman, Pollack, & Burton (2008) use course level data in predicting grades in different disciplines.

Program 2000-2001 states that the Verbal SAT and Math SAT have a correlation of 0.47 and 0.48 respectively with FGPA. According to Borenstein and Cohen (1989), these sets of correlation are medium, not large nor small. Some other studies found the correlation hovers around 0.5 regardless of the number of colleges studied (Fincher 1974; Ford and Campos, 1977; Dittmar, 1977; Goldman & Hewitt, 1976; Goldman & Richards, 2005).

Some researchers point out that, while the SAT is predictive for white male students, it is not predictive for minority and female students' college grades. Fleming (2002) showed that SAT scores did not correlate significantly at all with college GPA among black students at New York University. Moffatt (1993) conducted a study of 570 undergraduate students in a regionally accredited Southern college and found that the SAT is biased against black students regardless of age. One possible reason for this finding is that the SAT scores gap does not translate into a college GPA gap by demographic groups. Bowen and Bok (1998) found that average SAT composite scores for white applicants was 186 points higher than that of the African American applicants among the 1989 applicants in highly selective universities. Almost 75% of white applicants scored above 1200 in their SAT while only 29% of black applicants did. Using the data from the Florida State University System of over one million applicants, Micceri (2007) reported that whites consistently outscore minorities by an average 60 points, while males held an average advantage of 75 points over females with the same high school GPA. Five possible reasons for these findings are: (1) the SAT is biased, (2) whites are inherently smarter, (3) wealthy white students have better preparation for the test, (4) grades are biased in college, and/or (5) affirmative action exists in college admission, in which case individual colleges have a lower SAT threshold for minority students.

Besides different measures for predictive ability and the issue of the SAT's possible bias, omitted variable bias may distort the SAT's prediction. Rothstein (2004) points out that many studies omitted students' background characteristics, which inflates the SAT's apparent contribution. Using data from University of California administrative records (22526 California residents from the high school class of 1993), Rothstein found that the predictive ability of the total SAT scores on college success falls by 20% in the characteristics-adjusted model. The results showed that, in order to avoid omitted variable bias, researchers must consider students' background. This paper suggested that omitted course information may also cause omitted variable bias. A component of GPA, grading difference across professors and departments, may distort the meaning of GPA. The SAT may overestimate a student's GPA in a challenging class or major and underestimate a student's GPA in a less demanding class or major.

The only two studies that look at the SAT's predictive ability across courses and course type are from the College Board's Data, Reports and Research. Using data from 26 colleges, Bridgeman, Pollack and Burton (2008) found that the correlations of the SAT composite scores with cumulative GPA for Education courses, English courses, and Social Science, Science, Math & Engineering courses are 0.35, 0.48, and 0.54 respectively after taking into account high school GPA (HSGPA). Bridgeman and Wendler (1989) also look at the predictive ability of the SAT in predicting college Mathematics courses with data from 10 colleges. They found that the predictive ability of the Math SAT is around 0.35 for all math courses. The results from these two studies suggest that the prediction of the SAT for college GPA is more similar within departments and different between departments. One possible reason is the grading difference between departments, which causes the sorting effect. This paper looks at specifically whether this sorting effect has an impact on the SAT's prediction of college GPA. The first part of the

analysis looks at the prediction consistency for different majors. Since none of the prior studies have taken into account both the within and between-course variation, the second part of the analysis will also investigate how the SAT's prediction alters after controlling for specific course variables.

III. Theory

The goal of the Macalester College Admissions Office is to select applicants with excellent academic success (most importantly) and varied personal characteristics, with the goal of creating an intellectually gifted and interesting freshman class. As the agent in admission decisions, the admissions office wants to select the most suited applicants with applicants-reported information. With the difficulty in quantifying personal characteristics and prohibited access to personal statement for non-admission officers, this paper focuses on the goal of academic strength rather than personality.

Admission offices mainly use the SAT, high school grades (HSGPA), and strength of high school curriculum to predict students' academic strengths. Since different high school curriculums use various grading systems and criteria, the SAT acts as the only universal common yardstick to measure academic strength. Figure 1 shows the ideal positive relationship between the SAT and grades. Previous research has largely investigated how well the SAT predicts college grades, which is the slope of figure 1. What researchers have overlooked is the effect sorting in college has on the prediction of the SAT. Each department has its own grading criteria and standard; for example, the science discipline is more objective, while the humanities discipline is more subjective; the Mathematics department values more quantitative skill, while

the Art department emphasizes creativity and hands-on skill. Even within departments, professors may also grade differently depending on their standards.

Students also have different preferences when self-selecting themselves into classes. Their choice may be based on interest, academic strength, major requirement, or how tough or easy the grader is rumored to be. The relationship between SAT scores and grades within classes may be similar to figure 1 while the relationship would be closer to figure 2 without taking account for course variation. For example, the average letter grade for the Economics department in Macalester College is B, while that of the Education department is A-. The SAT would, therefore, predict differently for Economics and Education classes. In comparison to the SAT's prediction of a selective Economics class (Y1), Y2 illustrates the relationship between the SAT scores and Education class grades, given that the SAT scores ranges are similar. Y2 shifts up since the Education department gives higher grades *ceteris paribus*. As a result, the combination of all sorting effects would prevent the SAT from accurately predicting college success.

The assumptions of the model are as follows:

(1) The Macalester College Admissions office uses SAT scores in the admissions formula to help predict college success without using cut-off points.[5] If the admissions office uses cut-off points, the data sample would not have any information about students with SAT lower than the cut-off points. My estimation would contain selectivity bias.

(2) Three factors affect students' major or class choice: grades, utility or interest, and career goals. These factors lead students into specializations in a major or class, in which they would obtain the highest grade with lowest opportunity cost. With this assumption, GPA will be a true reflection of college success.

[5] An interview with an admissions officer has confirmed this assumption.

(3) Grading is unbiased. The GPA is only an accurate proxy of college success if professors grade according to students' performance within a department and/or course. It is possible, nevertheless, that professors grade according to how much they favor the students or based on other criteria.

(4) Students have equal access to education resources and assistance from faculty and staff members. Students' course grades and GPA are a reflection of their academic strength and major or course variation only. This assumption may not hold true based on students' family income inequality, e.g. if there is an inequity in access to computers, books, or paper. Macalester College has worked to avoid the problem by providing scholarships and free access to computers. Also, if some faculty or staff favor particular students, those students may benefit academically from spending more quality time with the professionals.

(5) Students with higher ability or academic drive will get higher grades, *ceteris paribus.*

Assumptions (2) to (5) ensure college GPA as a true reflection of academic success other than the reason of sorting—grading difference among professors and departments.

Assuming the admissions officers have the information of student *i*'s SAT scores and other characteristics, such as gender, nationality and leadership experience, the admissions office intends to use this information to predict academic outcome, y_i, represented as

$$y_i = f(\text{academic strength, non-academic characteristics}) \quad (1)$$

To measure the effect of sorting, I use real fixed effects to control for any grading or performance difference that is specific to the year, major, department, and course. This paper compares the prediction in three situations: (1) controlling for neither major nor course fixed effect; (2) controlling for major fixed effect, and (3) controlling for course fixed effect.

Cum GPA= f(academic strength, non-academic characteristics, major) (2)

Course grade= f(academic strength, non-academic characteristic, course info) (3)

I use equation (2) to test the accuracy of the SAT's prediction of cumulative GPA across different majors. If sorting does have an effect, the predictive power of the SAT will differ by major. I use cumulative GPA as the outcome variable. Grove, Wasserman, & Grodner (2006) have found that collegiate GPA data offer the best proxy for students' individual propensities to learn in college. Cumulative GPA is also a more comprehensive measurement for student's academic strength than freshman grades. Equation 3 contains real fixed effects for the department, course number, and semester the course is offered.

IV. Summary Statistics

To examine whether sorting makes a difference for the SAT's ability to predict college GPA at Macalester College, I use data provided by Macalester College's Institutional Research Department. The data include the dependent variables of the estimation equation, Mac cumulative GPA and course grade, and explanatory variables, like the gender, high school GPA, SAT Mathematic scores, and SAT Verbal scores of all of the students of graduation classes 1996 to 2005. Data for students' home country/ state, and first, second & third major and minor are also available for graduating class 1996-2003.

The summary statistics draw attention to GPA, SAT scores, and high school GPA. The distribution of the GPA is skewed to the left. The mean GPA of Macalester students from all graduation classes is 3.38 (range from 1.64 to 4) with a median of 3.44. Table 1 shows the number of observations, Mac cumulative GPA, and standard deviation by year and category. Across all categories, the means are different from each other by less than 1 standard deviation.

The table illustrates some notable differences between groups. International students have higher GPAs than their American peers at graduation. Female students generally achieve higher GPAs than males. Students with more majors also seem to do better than those with fewer majors, suggesting that perhaps students with more majors perform better rather than get distracted or overwhelmed by the amount of planning and work that comes with more majors. On the other hand, better students may be more willing to take on a second major or more able to balance between work and extra-curricular activities.

Figure 3 summarizes the GPA by majors, department, higher and lower level courses.[6] Overall department mean is the average GPA of all courses in a department taken by both majors and non-majors. The Chemistry, Economics, and Mathematics departments give the lowest grades, while the Neuroscience, Legal Studies, and Music departments give the highest grades. Breaking down to different level courses, the lower-level course grades are very similar to the overall department mean GPA[7]. The course grades in higher-level courses seem to be much higher when compared to the overall department mean GPA and lower-level course average grade. The Economics, Hispanic Studies, and International Studies departments give the lowest grades to the higher-level course takers; on the other side of the spectrum, higher-level courses gets the highest grades from the Education, Women's, Gender & Sexuality Studies, and Neuroscience departments. Last but not least, the major mean is the average cumulative GPA of the department majors. The disparity of major mean seems a lot smaller than the other three, ranging from 3.24 to 3.49. Economics, Philosophy, and Communication Studies majors achieve the lowest GPAs while Classics, Hispanic Studies, and Women's Gender & Sexuality Studies majors achieve the highest grades.

[6] Figure 3 is discontinuous because some values are missing.
[7] Lower-level courses enroll mainly freshmen and sophomores, while higher-level courses enroll mainly juniors and seniors. Higher-level courses also tent to have fewer students than lower-level courses.

The GPA summary illustrates the effect of sorting. More selective departments, like Economics and Chemistry, sort students through natural selection. Only those students who do well will get better grades and be encouraged to major in those subjects. Students who end up in those majors also obtain a higher average grade than the lower level course takers. As a result, those selective departments have their major mean and higher-level course grade exceeding the lower lever course grade. It supports assumption 2 in the theory section, which states that students specialize in majors with the least opportunity cost.

Table 2 contains the summary statistics of the SAT scores, showing that the means of SAT Math and Verbal scores are approximately the same, with a higher standard deviation for the Verbal test. Like Mac cumulative GPA, SAT scores also exhibit a positive trend across time. Figure 4 shows the Math and Verbal SAT scores by major with the total SAT scores ascending from left to right. Most majors have higher verbal scores than Math scores with the exception of all science subjects, Economics, Asian Studies and Individually Designed majors. Mathematics majors have the highest Math scores; Hispanic Studies majors have the lowest Math, Verbal, and total scores. Comparing figure 3 with figure 4, many majors with higher test scores attain relatively lower GPAs, and vice versa. For example, Neuroscience majors have one of the highest GPAs, but the second lowest SAT scores.

It would be meaningless to interpret the average of high school GPA (HSGPA) as reported in the admissions files since the data range from 1.67 to 105. The range is due to the different grading systems across countries, states, school districts, and schools. My data does not have the information about individuals' high school curriculum. After researching, the most common grading systems are the 4 point GPA scale, 5 point AP scale, 12 point scale, 20 points

scale, 45 IB scale, and 100-point scale. I rescaled the GPA data into the 4-point GPA scale.[8] I dropped the two observations that have HSPGA higher than 100. After rescaling them, the mean is 3.52 with the maximum being 4 and minimum being 1.67.

V. Analysis

The analysis section contains two sections. Section V.1 investigates the sorting effect by majors and discipline through looking at the consistency of the SAT prediction for cumulative GPA. Session V.2 compares the overall SAT prediction for grades controlling for all sorting effects.

V.1 Sorting by Major and Disciplines

As reviewed in the literature, most papers use equation 4 as the basic model for looking at the predictive ability of the SAT (Crouse and Trusheim 1988; Rothstein 2004; Thomas 2004; etc):

$$y_i = \alpha_i + \beta SAT_i + \mu HSGPA_i + \gamma \mathbf{X}_i + \epsilon_i \quad (4)$$

Building on the guiding equations 2, equation 4 regresses outcome y_i on SAT scores and high school GPA (HSGPA) as proxies for high school academic success. $\mathbf{X}_i$ is a vector of student i's non-academic characteristics, including gender, graduation class, country origin/ state, and the number of concentrations.[9] Equation 5 builds upon equation 4 by also controlling for college major.

[8] I rescale anything between 4 to 5 according to the AP scale, 5 to 12 according to the 12 points scale, 12 to 20 according to 20 points scale, 30 to 45 according to the IB scales and 45 to 100 according to the 100 scale into the 4 point GPA scale. I choose to convert to the 4.0 GPA because most entries seem to be on this scale, and also it is more comparable with the College 4.0 GPA scale.

[9] Concentrations include both majors and minor. I assign 1 to students with one major and 0.5 to students with one minor.

$$\text{cumulative GPA}_i = \alpha_i + \beta \text{SAT}_i + \mu \text{HSGPA}_i + \omega \text{major}_i + \gamma \mathbf{X}_i + \epsilon_i \quad (5)$$

Equation 6 applies a double log function for all the numerical variables in equation 5, so that the coefficients are interpreted as elasticities. Equation 6 also adds an interaction term between SAT scores and major to examine the consistency of the SAT's predictive ability for cumulative GPA for each major, represented as:

$$\ln(\text{cumGPA}_i) = \alpha_i + \beta \ln(\text{SAT}_i) + \mu \ln(\text{HSGPA}_i) + \omega \text{major}_i + \gamma \mathbf{X}_i + \xi(\ln\text{SAT}*\text{major})_i + \epsilon_i \quad (6)$$

I regress equation 6 separately for Verbal and Math SAT to measure the predictive ability of each of the test scores.

If the SAT is indeed a common yardstick even for college applicants intending to have different majors, ξ would be the same for every field; on the other hand, if sorting makes a difference, ξ would be significantly different across majors. The sorting effect occurs as students self-select themselves into majors with different grading standards and valued skills.

Several estimation issues arise with the regression. First of all, since this model involves a significant number of interaction terms (41 majors), the regressions run into a multi-collinearity problem. The preliminary regression keeps all of the dummies since the consequence of imperfect multi-collinearity only enlarges the standard error and does not cause bias in the estimates. Furthermore, most of the interaction terms are significant. For a robustness check, I grouped the majors into science, humanities, and fine arts disciplines to alleviate the multicollinearity, similar results emerged. Secondly, the White test indicates heteroskedasticity in the data, likely resulting from the inconsistency in the number of students in different majors; therefore, this paper presents all results with robust standard errors. Third, I cannot observe the students who were not accepted into Macalester College. Applying to, getting accepted or rejected, and choosing a college are selection processes that contribute to a narrow SAT spectrum. Since accepted students probably have higher than average SAT scores, they will more

likely to earn a higher GPA in Macalester College if they choose to come. On the other hand, students with SAT scores in the bottom 15% percentile may not get accepted because they have a much lower chance to succeed at Macalester College. As a result, the data sample of admitted students only produces a lower correlation between the SAT scores and grades. Nevertheless, I cannot correct for the selection bias due to limited data.

The results in table 3 summarize the SAT's consistency in predicting cumulative GPA by majors, taking into account students' characteristics and graduation class. Column 1 and 2 show the regression results with Math and Verbal SAT scores respectively. Gender, high school GPA (HSGPA), and total concentration are all significant at the 99% level. Holding all things constant, female students attain 0.02 to 0.03% higher cumulative GPA than their male peers. A 10% increase in HSGPA translates into a 2.8% gain in cumulative GPA. Students with more concentrations also achieve higher GPAs. These three coefficients remain consistent for different specifications in this section.

To find the effect of the SAT on performance by major, I interact ln(SAT) with a dummy variable for major. By summing the coefficients on ln(SAT) and on the interaction term, the effect of the SAT can be interpreted. I also tested whether the sum of those coefficients was also statistically significant. Most of the coefficients are significantly different from zero. The SAT predicts best for Math majors' cumulative GPAs. A 10% increase in Math SAT scores translates into a 9.35% increase in cumulative GPA for mathematic majors with 99% significance. The second most predictive major for the SAT is Physics with a coefficient of 0.685, significant at the 95% level. Math scores appear to predict best for either science or social science majors with an emphasis on the use of mathematics or statistics. On the other hand, a 10% increase in the

Math SAT scores indicates a decrease in GPA of 4.9% for German majors, 4.8% for Japanese Studies majors, and 4.7% for American Studies majors.

The Verbal SAT predicts best for Hispanic Studies majors. A 10 % increase in Verbal scores translates into a 6.53% increase in cumulative GPA. The second is French with a coefficient of 0.426, followed by Religious Studies (0.4), Sociology (0.347), Computer Science (0.298), Communication Studies (0.286), and English (0.259). The results indicate that the Verbal SAT is more correlated with cumulative GPA for language and humanities majors. On the other hand, a 10% increase in Verbal SAT scores for American Studies correlates to an 8.93% decrease in GPA. A similar trend is observed for the Japanese Studies major, though with a smaller coefficient of -0.297. I do not have an explanation for these negative coefficients.

Figure 5 graphs the predictive ability with both significant and insignificant coefficients by majors. It is arranged with ascending predictive power of Math scores from the left to right. The consistency of the SAT's predictive ability varies from major to major. The predictive power of Math and Verbal scores tend to have the same coefficient sign within a major, for example both negative for Japanese studies, American Studies, and Dramatic Arts; both positive for Political Sciences, Philosophy, Environmental Studies, etc. Worthy of note is the disparity in the SAT's predictive ability occurs across majors, with some significantly different from each other.

For a robustness check, I grouped the majors into five different disciplines: Self-designed, Science, Social Science, Humanities and Fine Arts. I also recomputed the coefficients and their significance under the combined coefficients of *ln(SAT) and the interaction term, ln(SAT*discipline),* to show the SAT's predictive ability in each discipline. Math scores predict best for Science majors, which is consistent with the finding from the preliminary result. Verbal

scores predict best for the Humanities discipline, which is also consistent with the previous result. After aggregating majors into disciplines, the disparity of the SAT's prediction reduces. The result still shows that ξ is different across majors and disciplines. The sorting effect taking place here is likely due to the nonlinear relationship between the SAT and grading criteria. The SAT just cannot measure equally the skills valued by all majors. To further take account for the sorting effect resulting from grading difference with departments, course-level data are used in the following section in the studies of the SAT's predictive power.

V.2 Sorting by Course

While only a few papers in the SAT literature use course-level data, this paper argues that taking account of between-course variation is necessary. Not only do departments value different kinds of skills, but professors also grade differently within departments. Within a department, a professor may choose B as the average grade, while another professor may choose C instead. I use a fixed effect methodology to control for course, time, and students specific effects. Building on guiding equation 6, the endogenous variable of equation 7 becomes student *i*'s GPA in course number $_j$ in semester $_k$. The course dummy controls information about department, semester, and number for the course:

$$\ln(\text{course grade}_{ijk}) = \alpha_{ijk} + \beta\ln(SAT_{ijk}) + \mu\ln(HSGPA_{ijk}) + \lambda\,\text{course}_{ijk} + \gamma\mathbf{X}_{ijk} + \epsilon_{ijk} \quad (7)$$

Same as equation (6) , equation (7) includes both the SAT scores and HSGPA as proxies for academic characteristics and the vector X of students' non-academic characteristics. Estimation issues includes slight multicollinearity between ln(HSGPA) and both of the SAT scores, but since all of the estimated coefficients are significant, I took no action. A Hesttest also indicates heteroskedasticity in the data; therefore, I ran all regressions with robust standard errors.

Table 5 compares the results of the SAT's predictive ability for both cumulative GPA and course grades. After controlling for class variation, the coefficient of the Math scores in column 4 is 50% larger than that in column 1. These two coefficients are significantly different at the 99% level, meaning that sorting between and within departments makes a difference in the Math SAT's prediction of grades. Since the confidence levels of the SAT coefficients in column 2 and column 5 overlap, no evidence supports the theory that course sorting makes a difference in Verbal scores' prediction. The coefficient of the total SAT scores increases by more than two standard errors after controlling for class variation (see column 3 and 6). The results suggest that sorting does make a difference in studying the SAT's predictive ability for grades.

Controlling for course variation provides a fairer mean of evaluating the SAT's prediction of grades. The Math SAT is 70% more predictive than the Verbal SAT. One possible reason is that the Math SAT measures one's quantitative skills, which are very useful and important in many sciences and social science classes. Verbal, comprehension, and writing skills are also very important in all disciplines, so one would expect the Verbal SAT to have greater predictive power. Perhaps the Verbal SAT just does not test those skills properly. In any case, it appears that the Math SAT is more useful than the Verbal SAT in predicting grades.

Worth noting is that the Math SAT is still somewhat predictive of course grade after controlling for all student, time, and course effect. A 10% (120 points) increase in math SAT translates to around 3.1% (0.12 points) increase in GPA These coefficients are consistent with those estimated by Rothstein (2002). The results show that the SAT's predictive ability is reliable and valid. Comparing with the coefficient on high school GPA (HSGPA), HSGPA has dropped by at least 10 times after controlling course variation. The result rejects the conventional theory, which suggests HSGPA is the most predictive measure for college success. It is possible

that high school grades capture a similar effect of sorting, and the effect disappeared after controlling for college sorting effect.

VI. Conclusion

The SAT today serves as a powerful tool in education resource allocation. In college admission processes, the SAT indicates a student's potential to excel and benefit from the scarce education resources. To maximize the utility (productivity) of valuable college resources, the students with the highest potential for success would be given the best education opportunity. Given this important role, the SAT should be a reliable and valid measure for college readiness.

Despite the SAT's popularity, scholars still debate about its predictive ability and fairness. This paper brings in the idea of sorting, which has been largely overlooked in previous studies. Results from this study support the hypothesis that ignoring the sorting effect distorts the prediction of the SAT and causes omitted variable bias in studies that only look at freshman or cumulative GPA. This paper tackles three kinds of sorting effects: major, discipline, and course effects. The results by major and discipline shows that the SAT predicts GPA for various majors and disciplines differently with statistical significance. A general trend shows that the Math SAT scores are more predictive for science and math-related social science majors while the Verbal scores are more predictive to language majors. Despite this general trend, the Math and Verbal SATs are not equally predictive for their related majors. In addition, some majors' GPAs even have a negative relationship with the SAT scores. After controlling for course variation, the estimated coefficients on Math SAT and total SAT in this paper increase with statistical significance. This result implies that students sort themselves into different classes in which professors grade with different criteria even within departments. On the other hand, the

predictive power of the Math SAT is 70% higher than that of the Verbal SAT. This implies that the Verbal SAT may not be a good measure for college-readiness.

A possible explanation behind this result is that students sort themselves into different majors with various grading criteria and valued skills. Those skills are related to the SAT scores in non-linear ways; for example, the Math SAT is better at measuring the quantitative skills that Math majors need; nevertheless, none of the SAT sections can measure the creativity and musical training required of Music majors. The SAT is, however, still predictive with 99% statistical significance after taking into account all course, department, and semester effects. The SAT may not be a perfect indicator, but it is robust and does provide useful information for an admissions office. In comparison to the prediction of high school GPA, the SAT does a much better job in predicting college academic grades. As Paulos (2008) concluded in his new article, "the SAT should not be made into a fetish, but it neither should it be ignored."

This paper is the first contribution of its kind in the SAT prediction literature. Similar research should be conducted with different university types and student bodies. Further research should also examine whether college GPA is a true reflection of students' success given the speculation of grading inflation and bias (Germaine, 2005). In addition, there needs to be alternative measurement in order to compare the fairness and prediction of the SAT. Future research should also compare the SAT predictive ability to other standardized admissions test, such as the ACT. The SAT Reasoning Test has also included a new writing section in 2005. It would also be interesting to investigate whether the added section improves the overall predictive ability of the SAT.

Acknowledgements

The author thanks her excellent advisors Daniel Kaplan, Karine Moe, Raymond Robertson, and Sarah West for guidance that helped shape the paper. The author also wants to thanks Lowell Busman, Yina Chng, Carl de la Cova, Nancy Eastham, Barbara Fadirepo, Becky Graham, Jeanette Hughes, Xiaobo Ke, Tina Kruse, Ruthanne Kurth-Schai, Kai Hi Liu, Long Ting Liu, Kathleen Macosko, Fiorella Ormeno-Incio, Yee Wan Szeto, Tiong Tan, Vicky Tan, Han Tsoi, Anastasia Verdoljak, Lijuan Wang, and participants at the 2009 Economics Scholars Program, National Conference on Undergraduate Research, and the 2009 Economics Honor Defense for support and comments that greatly improve the paper.

References

Armstrong, W. B., & Carty, H. M. (2003). *Reconsidering the SAT-I for college admissions: Analysis of alternate predictors of college success.* Paper presented at the Annual Meeting of the American Educational Research Association, Chicago, IL

Borenstein, M., & Cohen, J. (1989). *Statistical power analysis*. Hillsdale, N.J.: Lawrence Erlbaum Associates.

Bowen, William G., and Derek Curtis Bok. (1998). *The shape of the river: Long-term consequences of considering race in college and university admissions*. Princeton, N.J.: Princeton University Press.

Bridgeman, B., & Wendler, C. (1989). *Prediction of Grades in College Mathematics Courses as a Component of the Placement Validity of SAT-Mathematics Scores. College Board Report No. 89-9*. New York:College Board Publications.

Bridgeman, B., Pollack, J., & Burton, N. (2008). Predicting grades in different types of college courses. *College Board Research Report, 1*. New York:College Board Publications.

Crouse, J., & Trusheim, D. (1988). *The case against the SAT*. Chicago: University of Chicago Press.

Dittmar, N. J. D. (1977). *A comparative investigation of the predictive validity of admissions criteria for Anglos, Blacks and Mexican-Americans*. Thesis (Ph. D.)--University of Texas at Austin, 1977.

Fincher, C. (1974). Is the SAT Worth Its Salt? An Evaluation of the Use of the Scholastic Aptitude Test in the University System of Georgia Over a Thirteen-Year Period. *Review of Educational Research*. 44 (3), 293-305.

Fleming, J. (2002). Who Will Succeed in College? When the SAT Predicts Black Students' Performance. *REVIEW OF HIGHER EDUCATION.* 25, 281-296.

Ford, S. F., & Campos, S. (1977). *Summary of validity data from the Admissions Testing Program Validity Study Service: Prepared for the Advisory Panel on the Scholastic Aptitude Test Score Decline*. Princeton, N.J.: The Board.

Germaine, M.-L., & Scandura, T. A. (2005). Grade Inflation And Student Individual Differences as Systematic Bias in Faculty Evaluations. *Journal of Instructional Psychology.* 32 (1), 58-67.

Goldman, R. D., & Hewitt, B. N. (1976). Predicting the success of black, chicano, oriental and white college students. *Journal of Educational Measurement, 13*(2), 107-117.

Goldman, R. D., & Hewitt, B. N. (1976). The Scholastic Aptitude Test "Explain" Why College Men Major In Science More Often Than College Women. *Journal of Counseling Psychology.* 23 (1), 50-54.

Goldman, R. D., & Richards, R. (2005). The SAT Prediction of Grades for Mexican-American Versus Anglo-American Students at the University of California, Riverside. *Journal of Educational Measurement* 11 (2), 129-135.

Gould, S. J. (1981). *The mismeasure of man*. New York: Norton.

Grove, W. A., Wasserman, T., & Grodner, A. (2006). Choosing a proxy for academic aptitude. *Journal of Economic Education, 37*(2), 131-147.

Lemann, N. (1999). *The Big Test: The Secret History of the American Meritocracy.* New York : Farrar, Straus and Giroux.

Micceri, T. (2007). *How we justify and maintain the white, male academic status quo through the use of biased college admissions requirements.* Florida, Office of Planning and Analysis, University of South Florida.

Moffatt, G. K. (1993, February). *The validity of the SAT as a predictor of grade point average for nontraditional college students.* Paper presented at the Annual Meeting of the Eastern Educational Research Association, Clearwater Beach, FL.

National Association for College Admission Counseling. (2008). *Report of the commission on the use of standardized tests in undergraduate admission* National Association for College Admission Counseling. Alexandria, VA

Page, E., B., & Feifs, H. (1985). SAT scores and american states: Seeking for useful meaning. *Journal of Educational Measurement, 22*(4), 305-12.

Paulos, J. A. (2008, July 1, 2008). Do SAT scores really predict success? *ABC News.* Retrieved April 13, 2009, from http://abcnews.go.com/Technology/WhosCounting/Story?id=98373&page=1

Pope, J. (2008, June 17, 2008). Study: The new SAT is not much better. *Fox New*. Retrieved April 13, 2009, from http://www.foxnews.com/wires/2008Jun17/0,4670,NewSAT,00.html

Rothstein, J. M. (2004). College performance predictions and the SAT. *Journal of Econometrics, 121*(1-2), 297-317.

Rimer, S. (2008a, Sept 21, 2008). College panel calls for less focus on SATs. *The New York Times.* Retrieved April 13, 2009, from http://www.nytimes.com/2008/09/22/education/22admissions.html?ei=5070&emc=eta1

Rimer, S. (2008b, Sep 28, 2008). Study of standardized admissions tests is big draw at college conference. *The New York Times*. Retrieved April 13, 2009, from http://www.nytimes.com/2008/09/29/education/29admissions.html?ei=5070&emc=eta1

Stecklow, S. (1995). Cheat sheets: Colleges inflate SATs and graduation rates in popular guidebooks: Schools say they must fib to US news and others to compete effectively: Moody's requires the truth. *The Wall Street Journal*, , 1.

Thomas, M. K. (2004). Where college-bound students send their SAT scores: Does race matter? *Social Science Quarterly, 85*(5), 1374-1389.

University of California - San Diego (2009). SAT, Classroom Grades And College Readiness: Reflections On 100 Years Of Testing. *ScienceDaily*. Retrieved April 29, 2009, from http://www.sciencedaily.com/releases/2009/04/090414102650.htm

Weitzman, R. A. (1982). The prediction of college achievement by the scholastic aptitude test and the high school record. *Journal of Educational Measurement, 19*(3), 179-191.

Yablon, M. (2001). Test flight: The scam behind SAT bashing. *New Republic*, 24-25.

Zwick, R., & Sklar, J. C. (2005). Predicting college grades and degree completion using high school grades and SAT scores: The role of student ethnicity and first language. *American Educational Research Journal, 42*(3), 439-464.

Tables & Graphs

Table 1: Mac GPA by Category			
Category	Obs	GPA mean	GPA SD
Total	4164	3.389	0.350
Citizen Country/ State			
International	500	3.446	0.355
American	3607	3.392	0.344
Not listed	57	3.091	0.612
Gender			
Female	2398	3.42	0.315
Male	1420	3.38	0.381
Number of majors			
1	2367	3.353	0.366
2	932	3.465	0.302
3	48	3.570	0.294
Graduation Class			
1994	401	3.344	0.384
1995	416	3.347	0.372
1996	387	3.357	0.366
1997	391	3.388	0.340
1998	435	3.391	0.339
1999	408	3.388	0.319
2000	399	3.394	0.356
2001	431	3.416	0.329
2002	432	3.405	0.350
2003	464	3.440	0.334

Table 2. The SAT scores summary				
	min math	mean math	min verbal	mean verbal
1998	430	628	250	588
1999	380	638	300	602
2000	460	638	320	653
2001	400	639	340	651
2002	450	655	420	670
2003	480	650	400	675
2004	410	655	390	669
2005	470	661	330	676

Table 3. SAT's consistency in predicting cumulative GPA by Majors				
Dependent variable: log(cum GPA)	(1) Math		(2) Verbal	
Gender	0.0292***	(0.006)	0.0225***	(0.006)
ln(HSGPA)	0.282***	(0.039)	0.284***	(0.039)
Total concentration	0.0123***	(0.002)	0.0136***	(0.002)
ln(Math SAT) / ln(Verbal SAT)	0.188***	(0.026)	0.173***	(0.023)
ln(SAT) + interaction term with ln(SAT)				
American studies	-0.476***	(0.089)	-0.893***	(0.044)
Anthropology	0.232*	(0.135)	0.208**	(0.088)
Art	0.122	(0.845)	0.045	(0.082)
Asian Studies	0.506*	(0.310)	-0.07	(0.112)
Biology	0.284***	(0.064)	0.201***	(0.067)
Chemistry	0.285	(0.239)	0.040	(0.135)
Classics	-0.088	(0.138)	0.271**	(0.122)
Communication Studies	0.241***	(0.101)	0.286***	(0.082)
Computer Science	0.331***	(0.161)	0.298**	(0.138)
Dramatic Arts	-0.317	(0.204)	-0.210	(0.329)
Economics	0.357***	(0.106)	0.155***	(0.055)
English	0.151	(0.109)	0.259***	(0.096)
Environmental Studies	0.095	(0.112)	0.195	(0.129)
French	0.263*	(0.140)	0.426***	(0.044)
Geography	0.316***	(0.140)	0.314**	(0.130)
Geology	0.178	(0.132)	-0.005	(0.157)
German	-0.487*	(0.287)	0.148	(0.135)
Hispanic Studies	0.282***	(0.038)	0.653***	(0.084)
History	0.177*	(0.038)	0.134	(0.089)
Humanities and Cultural St.	-0.026	(0.101)	0.221	(0.172)
Individually Designed	0.199**	(0.090)	0.114	(0.170)
International Studies	0.218*	(0.125)	0.195**	(0.080)
Japan Studies	-0.484***	(0.088)	-0.297***	(0.049)
Latin American Studies	0.200	(0.256)	-0.021	(0.110)
Linguistics	-0.284	(0.389)	0.023	(0.317)
Mathematics	0.935***	(0.210)	0.321**	(0.137)
Music	0.219*	(0.120)	0.023	(0.079)
Neuroscience	0.277	(0.404)	0.262	(0.171)
Philosophy	0.087	(0.122)	0.237	(0.166)
Physics	0.685**	(0.346)	-0.160	(0.356)
Political Science	0.072	(0.103)	0.049	(0.102)
Psychology	0.132	(0.099)	0.102*	(0.058)
Religious Studies	0.222	(0.165)	0.400***	(0.154)
Russian	0.161	(0.242)	0.238**	(0.111)
Sociology	0.433***	(0.151)	0.347***	(0.117)
Spanish	0.310***	(0.117)	0.279**	(0.134)
Theater	0.311	(0.812)	-0.305**	(0.155)
Urban Studies	0.075	(0.156)	0.094	(0.122)
Women's and Gender Studies	0.282*	(0.157)	0.255***	(0.090)
Constant	3.857***	(0.587)	6.561***	(0.250)
Observations	1539		1539	
R-squared	0.229		0.226	

Notes: Robust standard errors in parentheses *** significant at 1% level, ** significant at 5% level, * significant 10% level The gender dummy assigns 0 for male and 1 for female.
Under *ln(SAT) + interaction term with ln(SAT)*, the coefficients and significance of the coefficients are determined using the lincom command in stata after add the coefficients of overall SAT and the interaction term.

Table 4. Predictive Power of Math and Verbal SAT for Mac GPA by discipline				
	(1) Math SAT		(2) Verbal SAT	
Gender	0.0298***	(0.005)	0.0187***	(0.005)
ln(HSGPA)	0.277***	(0.038)	0.276***	(0.039)
Total concentration	0.0125***	(0.002)	0.0143***	(0.002)
ln(SAT)	0.195***	(0.025)	0.174***	(0.022)
Discipline Dummy				
Science	-0.68	(0.643)	-0.63	(1.091)
Social Science	0.000265	(0.642)	-0.253	(1.072)
Humanities	0.194	(0.625)	-0.746	(1.079)
Fine Art	1.031	(0.748)	0.543	(1.146)
ln(SAT) + interaction term with ln(SAT)				
Self-design* ln(sat)	0.203**	(0.089)	0.112	(0.165)
Science*ln(sat)	0.302***	(0.047)	0.206***	(0.047)
Social Science*ln(sat)	0.199***	(0.046)	0.148***	(0.031)
Humanities*ln(sat)	0.168***	(0.041)	0.223***	(0.039)
Fine Art*ln(sat)	0.04	(0.076)	0.026	(0.072)
Constant	-0.466	(0.569)	0.117	(1.054)
Observations	1539		1539	
R-squared	0.187		0.190	

Notes: Robust standard errors in parentheses *** significant at 1% level, ** significant at 5% level, * significant 10% level
Under *ln(SAT) + interaction term with ln(SAT)*, the coefficients and significance of the coefficients are determined using the lincom command in stata after add the coefficients of overall SAT and the interaction term.

Table 5. The SAT's Predictive ability for cumulative GPA and course grade						
	Dependent variable: cumulative GPA			Dependent variable: course grade		
	(1)	(2)	(3)	(4)	(5)	(6)
ln(Math SAT)	0.194***			0.305***		
	(0.0252)			(0.0083)		
ln(Verbal SAT)		0.172***			0.181***	
		(0.0223)			(0.0069)	
ln(total SAT)			0.268***			0.359***
			(0.0287)			(0.0093)
sex	0.0294***	0.0175***	0.0252***	0.0402***	0.0269***	0.0338***
	(0.00542)	(0.00523)	(0.00522)	(0.0018)	(0.0018)	(0.0018)
ln(hsgpa)	0.273***	0.287***	0.262***	0.0383***	0.0479***	0.0435***
	(0.0383)	(0.0395)	(0.0383)	(0.0019)	(0.0020)	(0.0019)
indept				0.00572***	0.00563***	0.00572***
				(0.0002)	(0.0002)	(0.0002)
Control for	Graduation year	Graduation year	Graduation year	Class variation	Class variation	Class variation
Constant	-0.409**	-0.274*	-1.056***	-0.875***	-0.0798*	-1.479***
	(0.159)	(0.145)	(0.199)	(0.0536)	(0.0450)	(0.0671)
Observations	1539	1539	1539	64698	64698	64698
R-squared	0.159	0.159	0.178	0.285	0.275	0.287

Notes: Robust standard errors in parentheses *** significant at 1% level, ** significant at 5% level, * significant 10% level
The variable *indept* equal to the numbers of classes $student_i$ had taken from the same department the course grade was given.

Fig 1: relationship between SAT and grade without the effect of sorting

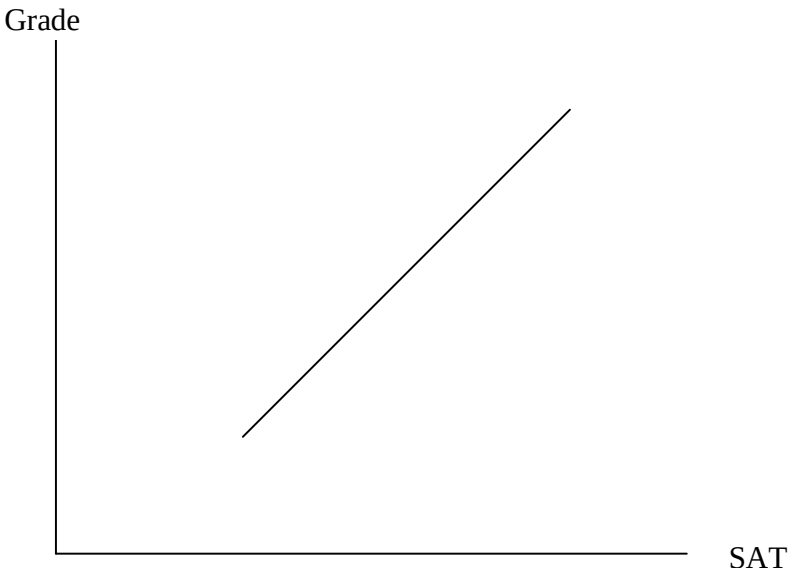

Fig 2: relationship between SAT and grade with the effect of sorting

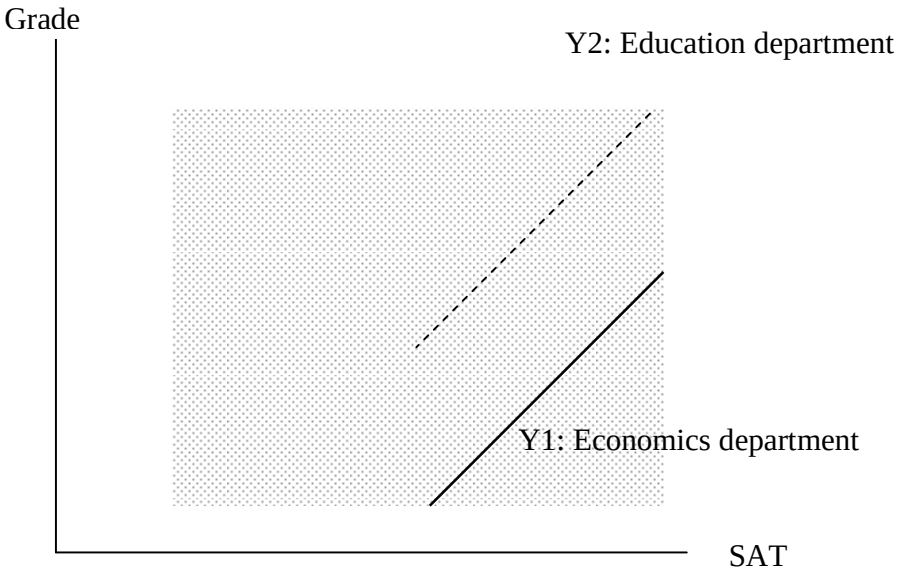

Fig. 3: GPA Summary

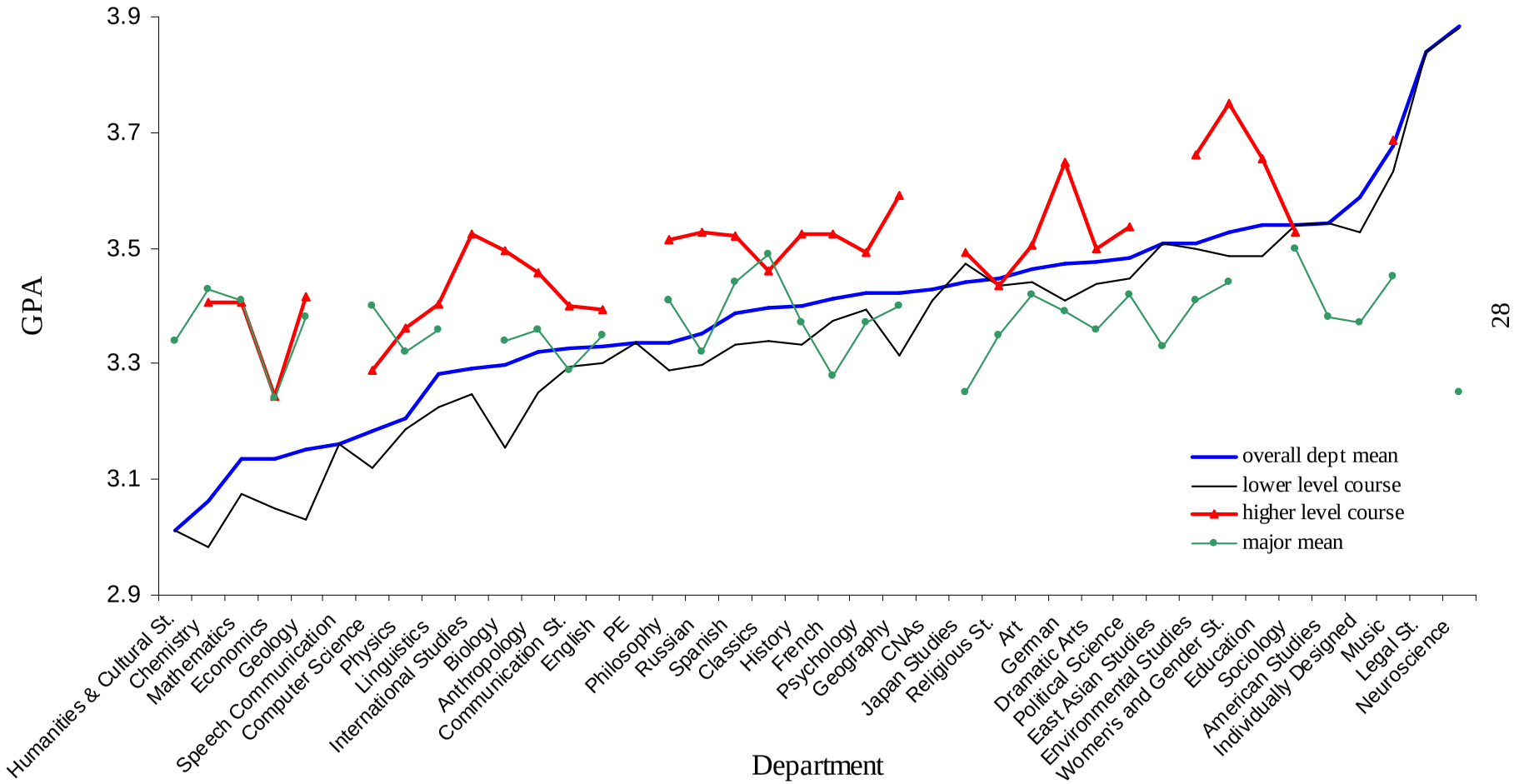

Fig. 4: Majors' Math & Verbal SAT scores

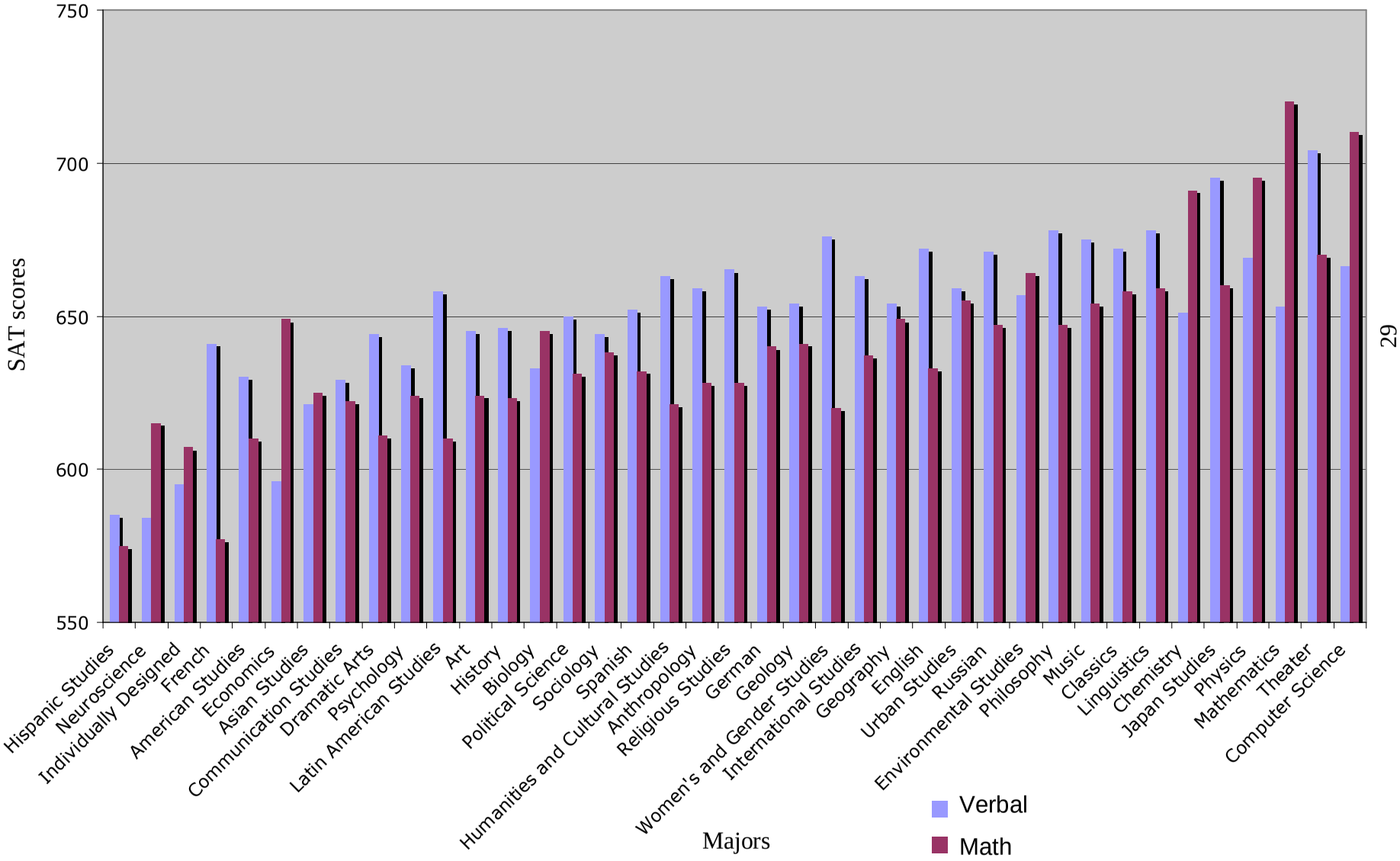

Fig. 5: The SATs predictive ability by Major