Predicting Grades in College Courses:

A Comparison of Multiple Regression and Percent Succeeding Approaches

BRENT BRIDGEMAN is a principal research scientist in the Center for Validity Research at the Educational Testing Service. His research focuses on diverse aspects of test validity and methods of communicating validity results.

JUDITH POLLACK recently retired from the Educational Testing Service. She was a senior data analyst specializing in the creation and analysis of large-scale databases.

NANCY BURTON was a senior research scientist in the Center for Validity Research at the Educational Testing Service. Her research focused on validation of high-stakes admission tests and investigating gender differences on high-stakes tests.

Author Note

We thank the College Board for supporting this research. Points of view or opinions stated do not necessarily represent official College Board position or policy.

Abstract

Two methods of showing the ability of high school grades (high school grade point averages) and SAT scores to predict cumulative grades in different types of college courses were evaluated in a sample of 26 colleges. Each college contributed data from three cohorts of entering freshmen, and each cohort was followed for at least four years. Colleges were separated into four levels by average SAT scores. Grade point averages (GPAs) for four categories of courses (English, science/math/engineering [S/M/E], social science, and education) were computed. Correlations of the combined predictors with course grades over four or more years, corrected for range restriction, ranged from .45 for education courses to .64 for S/M/E courses, and the SAT increment ranged from .03 in education courses to .08 in S/M/E courses.

Because these seemingly tiny incremental validities are frequently misinterpreted, a graphical display was developed that shows how the percent of students succeeding at a high level (cumulative GPA of 3.5 or higher) increases as SAT scores increase for students with similar high school grades.

Large-scale studies, combining data over many colleges and universities, have been used to demonstrate the validity of the SAT, or its predecessor the Scholastic Aptitude Test, for predicting college freshmen GPAs (e.g., Ramist, Lewis, & McCamley-Jenkins, 1994; Bridgeman, McCamley-Jenkins, & Ervin, 2000; Hezlett, Kuncel, Vey, Ahart, Ones, Campbell, & Camara, 2001; Geiser & Studley, 2001) or cumulative GPAs over a college career (Burton and Ramist, 2001). Typically, such studies use multiple regression methods to show the improvement in prediction when SAT scores are added to the high school GPA (HSGPA). The improvement is frequently described in terms of the additional variance in college grades that can be "explained" by the test scores. The additional explained variance attributable to the test is typically less than 10 percent. Testing critics suggest that such an apparently miniscule improvement is not worth the bother. The National Center for Fair and Open Testing (Fairtest) simply states, "the SAT I has little value in predicting future college performance" (Fairtest, 2003). Kidder and Rosner (2002) question the value of the SAT, citing data from a University of California study showing that "the SAT only adds 5.4% to the variance explained by HSGPA alone" (p. 193).

Unfortunately, "explained variance" is a very difficult concept to understand, and multiple regression may be obscure to the layman interested in testing issues. Nor is this a problem only for lay audiences, as even trained social scientists may severely underestimate the practical importance of apparently small amounts of explained variance (Rosenthal & Rubin, 1982). More recently, Wainer and Robinson (2003) reiterated the practical value of experimental treatments that explain only a miniscule percent of the variance in the outcome variable. This is most clearly seen if the success rates with and without treatment (or at different levels of a predictor), rather than the multiple correlation (R) or the squared multiple correlation (R^2), are used to demonstrate effectiveness. R^2 is an index of the percent of the variation in the criterion (such as college grade point average) that is "explained" by the predictors (such as test scores and high school grades). Wainer and Robinson (2003) cite data from a large-scale study in which 22,071 physicians were randomly assigned to take either aspirin or a placebo every other day over a five-year period, and the outcome variable was a heart attack. Using a traditional "explained variance" approach indicates that much less than one percent of the variance in getting a heart attack can be explained by taking (or not taking) aspirin. (The r^2 is .001.) Focusing instead on the number of people in each group who actually had heart attacks tells a far different story. In the group taking the aspirin there were 104 heart attacks, and in the placebo group there were 189, or almost twice as many.

This study seeks to show the value of grades and test scores for predicting success in college with a graphical display that requires the reader to understand only the meaning of percent. Thus, this study shows the percent of students in various HSGPA levels and test score categories who are successful in different types of college courses.

Method

Sample

The sample consisted of 26 colleges that agreed to participate in a multi-year validity study and submitted grades in individual courses for three cohorts of students who began college in 1995, 1996 and 1997. Each cohort was followed for at least four years. The sample of colleges was geographically diverse, included large and small public and private institutions, and represented a range of selectivity as indexed by average SAT scores. Differences between the yearly cohorts were trivial, so the three cohorts at each institution were combined. Most students in the cumulative GPA sample attended college for at least four years, and some attended for five or six years. Because students drop in and out of college, the eligibility for the cumulative GPA sample was not defined in terms of years attended, but rather by requiring grades for at least 24 courses. The sample of colleges were divided into four selectivity levels defined by average SAT scores (combined verbal and math). Level 1 colleges had scores around the national average of approximately 1000 while Level 4 colleges were highly selective with average scores over 1250. Although Level 3 consisted of only three institutions, all three were large flagship state universities and so were treated separately. Table 1 describes the sample.

Procedures

Colleges provided HSGPAs from their records and also provided grades in individual college courses. Cumulative GPAs were computed for students who attended college for at least four years and met the 24-course requirement. In addition to the overall GPA, separate GPAs were computed for English courses, physical and biological sciences/math/engineering (S/M/E) courses, social science courses, and education courses. A student had to have taken at least two courses in a subject area in order to be included in the GPA for that subject area.

The first step was to evaluate traditional multiple regression equations, predicting cumulative grades for the entire undergraduate career. As is customary in studies of this type, multivariate corrections for range restriction were made in order to estimate correlations in the potential applicant population, not just in the restricted group of students actually enrolled (Gulliksen, 1950, pp. 165-166). To apply this correction it was assumed that the potential applicant pool consisted of all students who had taken the SAT and thus used the national standard deviations and intercorrelations for SAT-V, SAT-M, and HSGPA (from the SAT Questionnaire) in the corrections. All correlations were computed separately within each institution, then the weighted average of these correlations across institutions at each college selectivity level was assessed.

For the alternative "percent succeeding" analyses, success was defined as achieving at least a 3.5 GPA in the subject. (Other criteria were also examined, such as 2.5 or above, but for the current illustrative purposes, the 3.5 criterion is sufficient.) Then the percent of students who were successful at five different levels of SAT scores was computed for students with comparable HSGPAs (either 3.3 to 3.7 or above 3.7). The five SAT levels, computed by combining verbal and math scores, were: 1 = 200-800; 2 = 810-1000; 3 = 1010-1200; 4 = 1210-1400; 5 = 1410-1600.

Results

Predicting Cumulative College GPA

Table 2 indicates the familiar finding that, overall, HSGPA is a slightly better predictor than the SAT, but this pattern is not repeated in all subgroups. In particular, the SAT appears to be the better predictor for males in all three minority groups and also for females in the Asian and African-American groups. Despite some minor variations, results were remarkably consistent across the colleges in the four different selectivity levels. As in previous studies, the incremental contribution of SAT scores over HSGPA appeared to be quite small, accounting for less than 10 percent of the variance in college grades ($.66^2$-$.59^2$ = .09).

Predicting Cumulative GPA in Particular Course Types

Table 3 shows the correlations predicting cumulative GPA in four course types. Given the similarities in correlations across colleges in different selectivity levels, only the combined results are presented here. Education course grades appear to be the most difficult to predict from either HSGPA or SAT scores, followed by English grades, with S/M/E and social science grades about equally predictable. The increase in predictability of college grades by adding SAT scores to predictions from HSGPA alone is greatest for the S/M/E courses.

Table 1. Participating Colleges By College Selectivity (Combined SAT Score), and Number of Students in Combined Cohorts

Level 1 (SAT < 1100)			Level 2 (SAT 1100-1199)		
Public/Private	**Region**	**Number**	**Public/Private**	**State**	**Number**
Public	Middle States	2,471	Public	New England	1,947
Public	Middle States	2,089	Public	Middle States	3,665
Public	Middle States	2,134	Public	Middle States	4,588
Public	Middle States	2,273	Public	West	3,687
Public	Middle States	2,115	Public	West	8,166
Public	South	1,326	Private	Middle States	3,675
Public	Southwest	1,975	Private	Midwest	1,251
Private	South	1,431	Private	West	979
Private	West	611	Private	West	620
			Private	Southwest	342
	Total	16,425		Total	29,262
Level 3 (SAT 1200-1249)			**Level 4 (SAT >1250)**		
Public	**South**	**8,280**	**Public**	**South**	**4,617**
Public	Midwest	8,232	Private	Middle States	2,620
Public	Southwest	14,256	Private	Midwest	717
			Private	Midwest	494
	Total	30,768		Total	8,448

Note.—"Number" indicates the number of students in combined cohorts with valid grades for the cumulative GPA analyses.

Percent of Students Succeeding by HSGPA and SAT Score Categories

Figure 1 shows the percent of students reaching a high level of success (a 3.5 GPA) after four years in college for students in two HSGPA bands and five SAT levels. Even within a relatively narrow HSGPA band (e.g., only students with a 3.7 average or higher), success rates differ dramatically across SAT levels. In the Level 1 and 2 colleges (top graph), fewer than 20 percent of the students with high HSGPAs, but SAT scores in level 2 (810-1000), were highly successful, but in the top SAT level, more than 70 percent were successful. Similarly, in the Level 3 and 4 colleges (bottom graph), fewer than 10 percent of the students with high HSGPAs were highly successful if they were in SAT level 2, but more than 60 percent were highly successful if their SAT scores were above 1400. The graphs also show that HSGPA matters; within a single SAT level, students in the high HSGPA category are more likely to be highly successful than students with HSGPAs between 3.3 and 3.7. Note that in order to present a valid and stable picture, we have omitted data points representing fewer than 150 students—for example, in the Level 1 and 2 colleges, only 49 students had combined SAT scores of 800 or lower AND HSGPAs higher than 3.7. In the middle three SAT levels, each data point represents the performance of thousands of students.

Figure 2 shows the success percentages by type of course for students in the 3.7 and above HSGPA category. For all types of courses, the likelihood of success improves dramatically from the lowest SAT levels to the highest within a given HSGPA category. Figure 2 also shows the differences in grading standards across different types of courses. Within a given HSGPA and SAT level, success percentages are highest for education courses and lowest for S/M/E courses.

Percent with High Success Level by Parent Education Category

This approach focusing on the percent of successful students is also useful for showing the relationship of parental education level and SAT scores to success in college. Three levels of parental education were created based on responses from the SAT Questionnaire. The first level indicates that neither

parent had completed college, the second level indicates that at least one parent had a college degree, and the third level indicates that at least one parent had a graduate degree. Focusing first on the columns on the right of Table 4 (number of students), the relationship between SAT scores and parental education is clear. At SAT Level 1 (in Levels 1 and 2 colleges), there are more than four times as many students whose parents had no college degree as had a graduate degree, but at SAT Levels 4 and 5 there are more students whose parents had a graduate degree than students whose parents did not graduate from college. Nevertheless, there are still thousands of students in the sample with SAT scores in Levels 4 and 5 whose parents did not graduate from college.

The left side of the table indicates that there is some relationship of parental education to this high level of accomplishment in college, but the relationship is not particularly strong when compared with the differences by SAT score level within the parental education category. For example, note that 31 percent of the students who were at SAT Level 4, and whose parents did not graduate from college, were highly successful. Going down just one SAT level, but focusing on students with at least one parent with a graduate degree, the success level drops to 17 percent. Clearly, the higher SAT score is a better indicator of likely success than is the parental education level.

Discussion

This study replicated previous research (Bridgeman, McCamley, and Ervin, 2000; Bridgeman et al., 2004; Burton and Ramist, 2001; Elliott and Strenta, 1988; Hezlett et al., 2001; Linn, 1982; Ramist, Lewis, and McCamley-Jenkins, 1994; Willingham, 1985; Wilson, 1983; Young, 1991a, 1991b) that shows a small incremental validity for the SAT in terms of multiple regression coefficients. The study also extended previous findings to predictions of grades in specific types of courses that were earned in four or more years of college study. More importantly, these effects are much more clearly seen in differences in the percent of students who are highly

Table 2. Prediction of Cumulative GPA, Corrected for Range Restriction

Gender and Ethnic/race Group												
Predictor	College Level	Gender Total		Asian		African American		Hispanic		White		Total
		M	F	M	F	M	F	M	F	M	F	M+F
HSGPA	1	.55	.59	.46	.57	.41	.48	.54	.40	.56	.59	.59
	2	.61	.60	.58	.54	.42	.42	.50	.59	.62	.62	.61
	3	.54	.59	.50	.57	.48	.51	.45	.54	.54	.60	.58
	4	.58	.58	.50	.45	.48	.39	.49	.32	.60	.61	.58
	Total	.57	.59	.53	.55	.45	.47	.47	.52	.58	.61	.59
SAT-V+SAT-M	1	.52	.59	.57	.61	.49	.57	.60	.39	.50	.57	.54
	2	.57	.59	.57	.59	.50	.43	.46	.60	.57	.59	.55
	3	.58	.65	.58	.64	.51	.55	.51	.61	.56	.59	.59
	4	.54	.55	.46	.39	.50	.45	.53	.54	.55	.63	.52
	Total	.56	.61	.56	.59	.50	.52	.51	.58	.55	.57	.56
HSGPA+SAT-V+SAT-M	1	.61	.68	.64	.73	.57	.63	.66	.48	.60	.60	.64
	2	.67	.68	.66	.65	.55	.49	.57	.68	.68	.69	.67
	3	.64	.71	.62	.69	.56	.61	.55	.66	.63	.70	.67
	4	.64	.64	.55	.49	.59	.52	.63	.58	.66	.68	.63
	Total	.65	.69	.62	.66	.57	.58	.58	.64	.64	.69	.66
SAT Increment	1	.06	.09	.18	.17	.16	.15	.11	.08	.05	.07	.06
	2	.07	.08	.08	.11	.13	.08	.08	.09	.06	.07	.06
	3	.10	.12	.12	.12	.08	.10	.10	.12	.09	.10	.09
	4	.06	.07	.05	.05	.11	.13	.14	.25	.06	.07	.05
	Total	.08	.10	.10	.11	.12	.11	.10	.12	.07	.08	.07
science, math, and engineering[a]												

successful rather than in terms of an increment in the multiple correlation. The SAT "explaining" less than 10 percent of the variance given HSGPA may seem trivial, but the difference between a 16 percent success rate and a 73 percent success rate for students with similar high school records, but different SAT scores, appears less trivial.

Although these graphs and tables could be used for making predictions about the likely success of individual students, there are other methods that are better suited for this purpose such as ordinary least squares regression for prediction of overall grades or logistic regression for predicting dichotomous success categories. But as powerful as these methods may be, they are not easy for admission professionals to explain to lay audiences. The intent was to show in a straightforward manner how preadmission measures are related to overall success in college and success in particular types of courses without having to explain probability estimates based on log odds ratios. Sophisticated analytical methods were purposely avoided that might produce more exact predictions but would be harder for an audience without statistical training to understand.

Table 3. Correlations and Multiple Correlations with Cumulative GPA in Four Course Types, Corrected for Range Restriction

Predictors	Course Types			
	English	S/M/E[a]	Social Science	Education
HSGPA	.51	.56	.57	.42
SAT-V	.45	.47	.51	.32
SAT-M	.42	.54	.48	.32
SAT-V + SAT-M	.48	.56	.54	.35
HSGPA+SAT	.57	.64	.63	.45
SAT Increment	.06	.08	.06	.03

Table 4. Percent of Students with Cumulative College GPAs about 3.5 by Parental Education and SAT Level in Level 1 and 2 Colleges

	Percent with GPA over 3.5			Number of Students		
SAT Level	Parent	College	Degree	Parent	College	Degree
	No	Yes	Grad.[a]	No	Yes	Grad.
1	2	1	0	516	161	92
2	6	6	6	4,932	2,709	1,805
3	15	16	17	7,993	6,645	5,817
4	31	33	34	2,616	3,524	3,751
5	55	65	64	173	330	502
Total	15	19	23	16,230	13,369	11,967
[a]graduate degree						

REFERENCES

Bridgeman, B., McCamley-Jenkins, L., & Ervin, N. (2000). *Predictions of freshman grade point average from the revised and recentered SAT I: Reasoning Test.* (College Board Report No. 2000-1). New York: College Board.

Bridgeman, B., Pollack, J., & Burton, N. (2004). *Understanding what SAT reasoning test scores add to high school grades: A straightforward approach.* (College Board Reseach Report No. 2004-4 and ETS RR-04-40). New York: College Entrance Examination Board.

Burton, N., & Ramist, L. (2001). *Predicting success in college: SAT studies of classes graduating since 1980.* (College Board RR-2001-2). New York: College Board.

Elliott, R., & Strenta, A. C. (1988). Effects of improving the reliability of the GPA on prediction generally and on comparative predictions for gender and race particularly. *Journal of Educational Measurement* 25(4), 333-47.

Fairtest (2003) SAT I: A Faulty Instrument For Predicting College Success. Retrived from http://fairtest.org/facts/satvalidity.html

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.* Retrieved from (http://www.ucop.edu/sas/research/researchandplanning/pdf/sat_study.pdf)

Gulliksen, H. (1950). *Test theory of mental tests.* New York: John Wiley and Sons.

Hezlett, S. A. Kuncel, N., Vey, M, Ahart, A., Ones, D., Campbell, J., & Camara, W. (2001). The effectiveness of the SAT in predicting success early and ate in college: A meta analysis. Paper presented at the AERA/NCME Annual Meeting, Seattle, Washington.

Kidder, W. C., & Rosner, J. (2002). How the SAT creates "built-in headwinds": An educational and legal analysis of disparate impact. *Santa Clara Law Review*, 43(1), 131-211.

Linn, R. L. (1982). Ability testing: Individual differences, prediction, and differential prediction. In A. K. Wigdor & W. B. Garner (Eds.), *Ability testing: Uses, consequences, controversies, Part II.* Washington, DC: National Academy Press.

Ramist, L., Lewis, C., & McCamley-Jenkins, L. (1994). *Student group differences in predicting college grades: Sex, language, and ethnic group* (College Board Report Number 93-1; ETS RR-94-27). New York: College Board.

Rosenthal, R., & Rubin, D. B. (1982). A simple, general purpose display of magnitude of experimental effect. *Journal of Educational Psychology*, 74, 166-169.

Wainer, H., & Robinson, D. H. (2003). Shaping up the practice of null hypothesis significance testing. *Educational Researcher*, 32, 22-30.

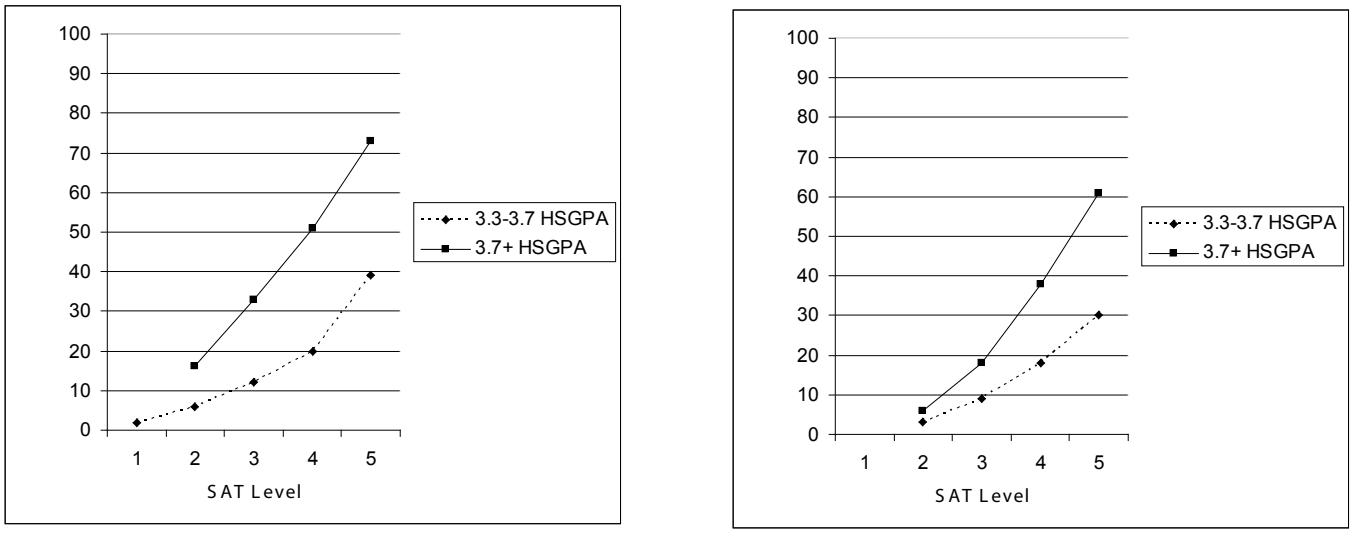
Willingham, W. W. (1985). *Success in college: The role of personal qualities and academic ability.* College Board.

Wilson, K. M. (1983). *A review of research on the prediction of academic performance after the freshman year* (College Board Report No. 83-2). College Board.

Young, J. W. (1991a). Gender bias in predicting college academic performance. *Journal of Educational Measurement* 28, 37-47.

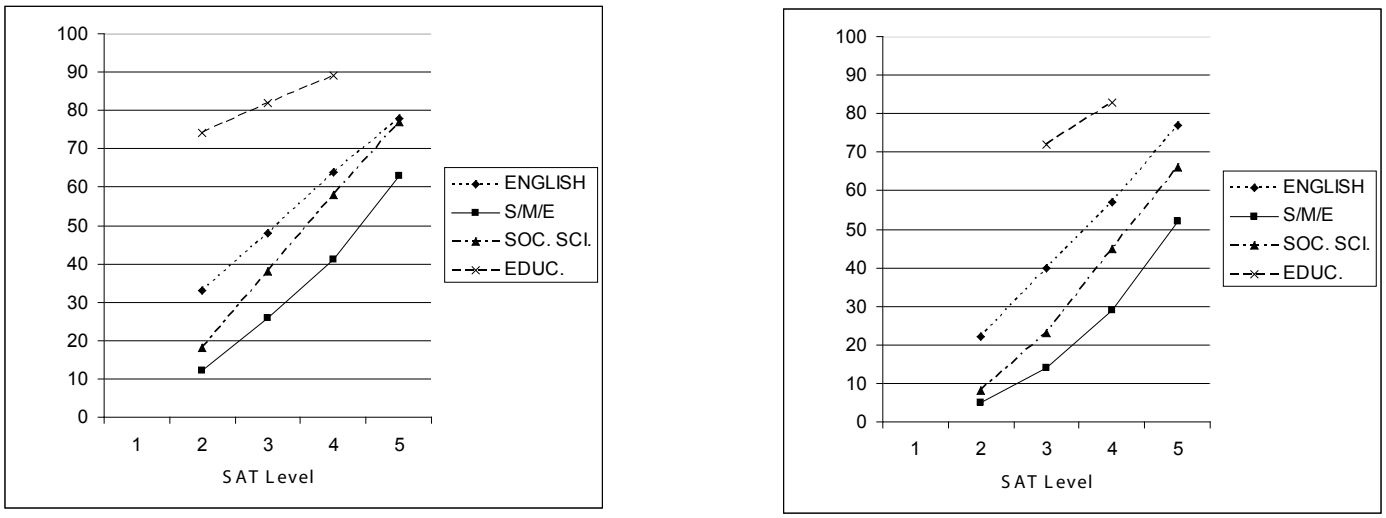
Young, J. W. (1991b). Improving the prediction of college performance of ethnic minorities using the IRT-based GPA. *Applied Measurement in Education* 4, 229-239.

FIGURE 1. For Level 1&2 colleges (left) and Level 3&4 colleges (right), percent of students with high school GPAs in the 3.3 to 3.7 range and 3.7+ range with cumulative GPAs of 3.5 or higher by SAT level.



Note. -SAT Levels: 1=200-800; 2=810-1000; 3=1010-1200; 4=1210-1400; 5=1410-1600. (Data points representing fewer than 150 students are omitted.

FIGURE 2. For Level 1&2 colleges (left) and Level 3&4 colleges (right), percent of students with high school GPAs in the 3.7+ range with cumulative subject GPAs of 3.5 or higher by SAT level.



Note. -SAT Levels: 1=200-800; 2=810-1000; 3=1010-1200; 4=1210-1400; 5=1410-1600. (Data points representing fewer than 150 students are omitted.