The College Board

Predicting Success in College: SAT® Studies of Classes Graduating Since 1980

Nancy W. Burton and Leonard Ramist

Predicting Success in College: SAT® Studies of Classes Graduating Since 1980

Nancy W. Burton and Leonard Ramist

Nancy W. Burton is senior research scientist, Center for Higher Education at Educational Testing Service.
Leonard Ramist is a retired ETS program administrator and a current ETS validity study consultant.

Contents

Abstract

Studies predicting success in college for students graduating since 1980 are reviewed. SAT® scores and high school records were the most common predictors, but a few studies of other predictors are included. The review establishes that SAT scores and high school records predict academic performance, nonacademic accomplishments, leadership in college, and postcollege income. The combination of high school records and SAT scores is consistently the best predictor. Academic preadmission measures contribute substantially to predicting academic success (grades, honors, acceptance and graduation from graduate or professional school); contribute moderately to predicting outcomes with both academic and nonacademic components (persistence and graduation); and make a small but significant contribution to predicting college leadership, college accomplishments (artistic, athletic, business), and postcollege income. A small number of studies of nonacademic predictors (high school accomplishments, attitudes, interests) establish their importance, particularly for predicting nonacademic success.

Key Words: Predictive validity, test fairness, SAT, admission testing, cumulative college grades, college graduation, high school record

I. Introduction

This review summarizes studies of the validity of the SAT and high school record as predictors of such long-term measures of success in college as cumulative grades, graduation, leadership, and postcollege income. Although most predictive validity studies use first-year grades as a proxy for success in college, long-term success in college is an equally important criterion to many colleges and universities. This review focuses on studies published after a previous review of cumulative college grade average by Wilson (1983), which covered classes graduating from college between 1930 and 1980. This review covers classes graduating between 1980 and the mid 1990s. To provide context, current long-term studies are compared to earlier studies predicting long-term criteria, and to both early and current studies predicting first-year grade average.

This report identifies those qualities of students that have proven to predict success in college. In an attempt to cover the outcomes important to different institutions of higher education, it presents as broad a picture as possible of the definitions of the success in college. Finally, it presents some ideas for future improvements by collecting information from a more varied array of admission credentials and criteria of success in college.

Over the years since the SAT was introduced, only a few institutions have studied its long-term validity. This review covers several validity criteria, but the two that were most frequently studied were cumulative GPAs and graduation. In a previous review of cumulative GPAs, Wilson (1983) reviewed studies of about 12,000 students who graduated from 40 institutions between 1930 and 1980. The portion of this review covering cumulative GPAs summarizes studies of about 80,000 students who graduated from 80 institutions since 1980 and one large study of students with disabilities who graduated from 124 different institutions. Selective institutions predominate as do small institutions. The portion of this review covering graduation includes even fewer studies—14 studies in all—but a number of them are based on large, multi-institution, representative samples.

This review tends to confirm the results of Wilson's (1983) review of studies predicting cumulative college GPAs. Both reviews found that SAT scores made a substantial contribution to predicting cumulative GPAs, and that the combination of SAT scores and high school records provided better predictions than either grades or test scores alone. A number of subgroups were studied in the current review period. Large studies were done for women students, African American students, and students with disabilities. The small studies reported for Asian American and Hispanic students should be considered as tentative support for the validity of the SAT and high school records. SAT scores and high school records predicted cumulative college GPAs significantly for each subgroup. The correlations found for subgroup students were large enough to be of practical use in admission, but the evidence is not definitive on whether the correlations were as large for subgroups as for the total group.

The relatively small number of studies of graduation were based on well over 100,000 students attending a broad sampling of nearly 1,000 colleges and universities. The results of these studies were consistent. The SAT and high school record both contributed substantially to predicting whether a student would graduate or not, but the correlation between predictors and graduation was smaller than the correlations between predictors and the cumulative college GPA. The probability of graduation was predicted as well for women, African American, Hispanic, Asian American, and Native American

students and students with disabilities as it was for the total group of students. Before presenting the review of long-term validity studies, we will provide some context by discussing important recent educational trends affecting admission and by discussing some technical difficulties that affect the interpretation of validity studies.

Recent trends in admission

Increasing numbers of Asian American, Hispanic, and African American students are flowing into higher education. At the same time, universities are attracting nontraditional students such as older students, international students, and home-schooled students. The primary and secondary educational communities have been going through significant reform, incorporating national, state, and district curriculum standards. In many states and districts, assessments are being used to evaluate attainment of standards; some of these assessments are performance based. These reforms are changing both the education students receive and the information available about student achievement. Colleges and universities need to know whether traditional admission measures adequately and fairly evaluate students with different backgrounds and students who present unusual credentials.

Litigation and legislation are changing the legal basis for admission to higher education. A common legal definition of affirmative action, developed since 1978 when the Supreme Court ruled on *Regents of the University of California vs. Bakke,* has been challenged in California and Washington through legislation and in the three *Hopwood* states—Texas, Missouri, and Louisiana—through litigation. In these states the effect has been to forbid the consideration of race, ethnicity, and gender as factors in admission.

Data are accumulating on the results of admitting undergraduate, graduate, and professional classes without affirmative action. Simulations (Bowen and Bok, 1998; Nickens, 1998; Wightman, 1997) show sharp decreases in admission of minority students if the current practice of basing admission decisions largely on previous grades and admission test scores is continued without considering race and ethnic group.

Actual enrollments in California and Texas showed declines in minority students despite such changes as the Texas policy of admitting all in-state students in the top 10 percent of their high school class, regardless of school quality (Cohen, 1998). The long-term effects are not yet known, although minority enrollments appeared to improve somewhat after the first year of implementation in Texas (Roser, January 19, 1998, September 15, 1998, December 28, 1998).

These events have caused some colleges and universities to question their current practices. Are the prevailing admission credentials good and fair predictors of important outcomes of a college education? Is there current information, based on the populations of students now attending colleges and universities? Are there other potential admission credentials that should be considered?

Before turning to the studies that may help answer these questions, it is necessary to review some of the technical problems that complicate interpretation of studies based on different populations of students, different courses of study, different admission policies, and different institutional missions.

Technical issues

Predictive validity studies are conducted to evaluate the admission process—the information used and the decisions made. The studies are typically based on a statistical correlation between admission credentials ("predictors") and available measures of success in college ("criteria"). The correlation depends as much on the criterion measure as the predictors. To be sure that low correlations indicate a problem in the admission process, one must eliminate the possibility that other aspects of the study are lowering the observed correlation coefficient. The major factors affecting interpretation of correlational studies (other than the quality of the predictors) include (1) restriction in the range of talents of the students taking a given college class (the cumulative result, perhaps, of decisions by both students and institutions that lead students into different colleges or universities and different courses of study); (2) professors using different grading standards; (3) other problems with measures of success in college, such as the unreliability of college grades. Success in college is a complex idea that the available criterion measures only partially and imperfectly approximate. All of these factors can affect the size of correlation coefficients regardless of the quality of the admission credentials used and admission decisions made.

Even with these problems it is still possible to establish basic facts such as the substantial contribution of SAT scores to predicting success in college. However, other important questions, such as whether SAT scores and high school records are equally fair for men and women, require a common basis for comparison that generally does not exist in raw validity data. For example, men tend to take more mathematics and science courses in college than women. Because these courses very frequently are more stringently graded than others, men on the average receive lower grades in college than

women. Prediction equations based on both men and women students will tend to predict higher grades for men than they actually receive (because they take more stringently graded courses than the overall average) and lower grades for women than they actually receive (because they take less stringently graded courses than the overall average). Unadjusted predictive validity data give the impression that prediction equations are biased against women since they receive higher grades than predicted. Predictive validity studies that adjust for the actual course-taking patterns of men and women reduce or eliminate this appearance of bias. In general, decisions that require the comparisons of different student groups or different admission measures may require adjustment to make the comparison fair.

Three of the important sources of error in unadjusted predictive validity studies are discussed below.

Restriction in the range of talent. As educational decisions are made, the range of talent in one university or major field or class may become quite different from others. It has been known for years that restriction of range mathematically lowers correlation coefficients. Ramist (1984) did a simple demonstration using data for 685 institutions in the College Board Validity Study Service. He searched for institutions with a full range of SAT scores and high school records. (Of the 21 colleges found, 18 had a religious affiliation.) In these institutions with an unrestricted range of students, the average correlation of SAT and high school record with the first-year GPA was .65, as compared to a median of .55 for all 685 institutions in the database.

Just as the range of talent in a college can become restricted, the choice of first-year courses will lead to some with a restricted pool of talent (advanced mathematics, for example) and some (such as required English composition) with a broad range of talent. Correlations of admission predictors with grades earned in first-year courses will be contaminated by these differences in restriction of range.

Willingham (1985) describes a process of students' migrating into majors with grading standards that best fit their level of preparation; that is, the best-prepared students tend to major in more stringently graded disciplines, and the least prepared students tend to go into more leniently graded disciplines. This finding suggests that restriction of range, which begins with college choice and continues with first-year course selection, continues to occur throughout the college years. The correlations of admission predictors with upper-division course grades will be reduced whenever restriction of range occurs. Restriction in range may partially explain the lower correlations of admission predictors with upper-division courses typically reported in the literature (Elliott and Strenta, 1988; Willingham, 1985; Wilson, 1983).

None of the studies of long-term success included in this review used existing statistical methods to correct for restriction of range. The reader can expect that these studies will underestimate the true validity of the admission credentials by an unknown amount.

Grading standards. Regression equations predicting the GPA in essence predict an average grade for individuals with particular values on the predictors. The prediction will not be correct for a student who takes mostly leniently graded or stringently graded courses—the predicted grade will be too low or too high. Ramist, Lewis, and McCamley (1990, p. 261) studied grading standards for first-year college grades and found that there was more than one grade point difference (on a 4-point grading scale) between the most leniently graded courses and the most stringently graded courses. The most leniently graded courses were physical education (actual grades .78 grade points higher than predicted), a combined group of classes including studio art, music, and theater (.56 grade points higher than predicted), and education (.50 grade points higher). The most stringently graded classes were a group of classes in science and engineering (.24 points lower), calculus (also .24 points lower), and biology courses for majors (.35 points lower). In a study of college grades, Elliott and Strenta (1988) found that controlling for grading standards increased the correlation of preadmission measures with the college GPA from .57 to .62 in the first year and from .41 to .51 in the senior year.

The lower uncorrected correlation found by Elliott and Strenta in the senior year (.41) as compared to the first year (.57) may be explained by Willingham's (1985) finding that able students tend to migrate into stringently graded majors while less-able students migrate into leniently graded majors, with the result that students with very different levels of accomplishment and knowledge receive similar GPAs in their major subject areas. This increased relativity of grading standards would lower observed correlations. (See also Goldman and Slaughter, 1976.)

Researchers have sought statistical methods to adjust for variations in grading (Braun and Szatrowski, 1984a, 1984b; Linn, 1966; Tucker, 1960). In the last decade, a growing number of researchers have proposed methods to adjust for variations among grades both within and across institutions. Elliott and colleagues at Dartmouth (Elliott and Strenta, 1988; Strenta and Elliott, 1987) matched students in pairs of courses to estimate and adjust course differences in grading stringency. Young (1990, 1991a, 1991b) used statistical equating theory to equate grading scales across general discipline areas.

Of the 174 institutions included in this review that studied the cumulative GPA, two adjusted for differences in grading stringency. In general, the reader can expect that the unadjusted studies will underestimate the true validity of admission predictors.

Other problems with measures of success in college. In different disciplines, different talents are useful and different class performances are rewarded. In performance areas such as art, music, and physical education, for example, the usual academic admission measures predict relatively poorly. Along with these systematic differences in grades there are also random differences. There is a long history of research on the inconsistency and subjectivity of teachers' grading practices, stretching back to Joseph Rice's turn-of-the-century studies of teachers' grades on spelling and mathematics exercises (Hillegas, 1912). These problems with systematic differences in the skills needed in different courses and unreliability of teacher-assigned grades are added to the problems caused by differences in the stringency of grading standards discussed in the previous section.

Although the grading literature will not be reviewed here, the reader can obtain an overview of the more recent literature by consulting Camara (1998), NCES (1984), OERI (1994), Robinson and Craver (1989), Willingham (1985), and Ziomek and Svec (1997).

Ramist and colleagues (Ramist et al., 1990; Ramist, Lewis, and McCamley-Jenkins, 1994) evaluated first-year grades in a broadly representative sample of 7,800 courses in 45 undergraduate institutions. They determined that the major sources of error in grades were restriction of range, variations in grading standards (including both differences in stringency addressed in the previous section and systematic differences in the skills required for different courses), and criterion unreliability. They developed methods of estimating the relationship of predictors to grades *within individual college courses* that controlled for all three sources of error. The resulting validity estimates are comparable across courses, disciplines, and institutions. The unadjusted correlation using the SAT and high school record to predict the first-year GPA in these 45 institutions was .48. After adjusting for restriction of range, grading standards, and criterion unreliability, the corrected correlation was .76.

When these methods were adapted to analyze the relationship between standardized achievement test scores and high school grades, Willingham, Pollack, and Lewis (2000) found that the corrected correlation between test scores and grades was .81 as compared to an unadjusted correlation of .62. This high school study, based on the national database of the National Educational Longitudinal Study, was able to go further than the typical admission study because measures of "studenting skills" reported by the students themselves (for example, "taking advanced electives") and their teachers (such as "does homework") were available. The final correlation between high school grades and test scores after controlling on studenting skills was .90.

The more complete methods of adjusting grades developed by Lewis, Ramist, and Willingham and their colleagues have not yet been applied to long-term criteria of success in college. The research available on grades over four years suggests that the problems of range restriction, differing grading standards, and criterion unreliability are likely to be at least as severe as they have proved to be for first-year grades. Thus, the studies of cumulative grades summarized in this review are likely to underestimate the true validity of admission procedures. Comparative data on unadjusted and adjusted correlations with first-year grades will be presented to enable the reader to estimate how much the validity of SAT scores and high school records for predicting cumulative grades may be underestimated.

Degree attainment is another major measure of success in college (Willingham, 1974). It is a measure of success in college that can be influenced by nonacademic factors such as persistence, efficient use of study time, and family support. Furthermore, a graduate-or-not (0/1) measure oversimplifies the process. Wilson (1978, 1980) demonstrated somewhat stronger correlations of predictors with a 7-point scale of levels of education reached from "returned for sophomore year" to "enrolled in graduate or professional school." Thus, the reader can also expect that the true validity of admission credentials to predict the completion of a course of study in college may be somewhat underestimated by studies employing a simple graduate/not graduate criterion.

Other important aspects of success, such as college accomplishments, college leadership and postcollege employment, income, and civic contributions are seldom measured with the important exception of two landmark studies by Willingham (1985) and Bowen and Bok (1998). These will be discussed in some detail.

Summary of technical issues. This brief review suggests that there are problems in both the predictors and criteria of success used in predictive validity studies. Grades, both high school grades used as predictors and college grades used as criteria, cover a broad range of academic and nonacademic skills, but they are known to have serious problems of comparability and reliability. Admission test scores are reliable and provide a common metric across all students, but they cover a limited range of academic skills and few or no nonacademic skills. The following review will indicate

that both predictors and criteria studied in the literature cover only a fraction of the qualities of successful college graduates.

This review of technical issues suggests that much of the difficulty in interpreting the predictive validity literature arises from inconsistent results because of varying degrees of range restriction, variations in grading standards, and unreliability of the criterion measure. Validity results may also be misinterpreted because all of these artifacts tend to depress observed correlation coefficients.

The studies reviewed below in general show moderate average correlations between common admission credentials and long-term measures of success in college, without corrections for the common artifacts that lower correlations. The results vary from study to study in part because different institutions and courses experience different levels of the various artifacts. In these less-than-perfect conditions, both SAT scores and high school records consistently make substantial contributions to prediction. Thus we believe that studies properly corrected for artifacts, particularly if the studies were conducted for a representative group of colleges and universities, would show a large contribution by SAT scores and high school records to predicting long-term success in college. Supplementing the current academic measures of success in college with some of the more important nonacademic measures of success and adding the appropriate nonacademic predictors—such as motivation, interests, or extracurricular accomplishments—would further improve the validity of admission decisions.

The following sections present the results of this review of long-term predictive validity studies for two measures that stand as proxies for success in college: cumulative grade average and graduation. They also review key studies of other measures of success. Comparisons are made to earlier studies and to the shorter-term criterion of first-year grade average. Comparisons are also made between unadjusted results and results adjusted for one or more of the problems just discussed. Results are presented for the total class and for such groups as men and women, race or ethnic groups, students with disabilities, and athletes.

II. Predicting Undergraduate Grades

This section reports the correlations of preadmission measures with cumulative undergraduate grade averages and compares results for cumulative grades with the results for first-year grades. The preadmission measures commonly used to predict success of undergraduates include the following:

- SAT I: Reasoning Test verbal (V) and mathematical (M) sections[1] (or ACT scores, which are not addressed here);

- an optimum combination of SAT V and M determined by individual institutions;

- a measure of high school record (usually cumulative GPA or rank in class);

- other predictors, such as SAT II: Subject Tests, accomplishments, motivation, or interests (seldom available); and

- multiple predictors. The most frequently available multiple correlation is the SAT I plus the high school record, but other multiples will be discussed where available.

Results

Table 1 displays studies of classes graduating from college between 1980 and the mid-1990s, including the number of institutions and number of students studied and correlations between predictors and the cumulative undergraduate GPA.[2] In addition to single predictors, several combinations of predictors are included. For the "best combination of SAT scores," multiple regression analysis was used to find the combination of verbal and mathematical scores that best predicted performance in each institution. Similarly, for "SAT + High School Record," multiple regression analysis was used to determine the best combination of SAT V, M, and high school record at each institution. A total of 30,000 students, graduating from 174 undergraduate institutions

---

[1]Prior to 1993–94, the College Board offered the Admissions Testing Program, which consisted of the Scholastic Aptitude Test (SAT) and a series of Achievement Tests. These were replaced by the SAT I: Reasoning Test and the SAT II: Subject Tests, respectively. In this paper SAT I and SAT II are sometimes used to refer to both the earlier tests and their replacements for consistency.

[2]In most studies, the cumulative GPA includes four or five years of grades earned through graduation. Some of the studies defined this criterion as the "final GPA," including students who officially withdrew as well as graduates. The GPA in this case can cover as little as one term or as much as six years. Still others used the "current GPA," which was the cumulative GPA for students at various stages of their undergraduate careers.

TABLE 1

Predicting Cumulative Undergraduate GPAs for Students Graduating Since 1980 (Uncorrected Correlations)

Predictor	Paper	Year	Number of: Institutions	Number of: Students	Group, if any	Correlation
SAT Verbal	Young and Barrett	1992	1	91		.17
	Shoemaker	1986	1	296	Engineering major	.21
	Shoemaker	1986	1	238	Computer science major	.23
	Crews	1993	1	336		.37
	Elliott and Strenta	1988	1	927		.38
	Moffatt	1993	1	28	Enrolled after age 30	.42
	Ra	1989	1	170		.42
	Young	1991a	1	1,564		.46
	Moffatt	1993	1	505	Enrolled before age 30	.50
	Total students			**4,155**	**Weighted average**	**.40**
SAT Math	Ra	1989	1	170		.28
	Crews	1993	1	336		.31
	Elliott and Strenta	1988	1	927		.34
	Moffatt	1993	1	28	Enrolled after age 30	.35
	Shoemaker	1986	1	238	Computer science major	.35
	Young and Barrett	1992	1	91		.41
	Shoemaker	1986	1	296	Engineering major	.43
	Young	1991a	1	1,564		.46
	Moffatt	1993	1	505	Enrolled before age 30	.49
	Total students			**4,155**	**Weighted average**	**.41**
SAT Verbal and SAT Math	Baron and Frank	1992	1	3,816		.20
	Nettles, Thoeny, and Gosman	1986	30	4,094		.31
	Moffatt	1993	1	28	Enrolled after age 30	.34
	Wolfe and Johnson	1995	1	201	Psychology class	.34
	Ra	1989	1	170		.39
	Tracey and Sedlacek	1985	1	1,339	White students only	.40
	Willingham	1985	9	3,442		.41
	Elliott and Strenta	1988	1	927		.43
	Ragosta, Braun, and Kaplan	1991	124	2,473	Students without disability	.52
	Moffatt	1993	1	505	Enrolled before age 30	.56
	Total students			**16,995**	**Weighted average**	**.36**
High School Record	Baron and Frank	1992	1	3,816		.30
	Young and Barrett	1992	1	91		.31
	Young	1991a	1	1,564		.35
	Wolfe and Johnson	1995	1	201	Psychology class	.40
	Elliott and Strenta	1988	1	927		.41
	Nettles, Thoeny, and Gosman	1986	30	4,094		.41
	Shoemaker	1986	1	238	Computer science major	.41
	Ra	1989	1	170		.44
	Willingham	1985	9	3,442		.45
	Leonard and Jiang	1995	1	10,000		.46
	Shoemaker	1986	1	296	Engineering major	.48
	Crews	1993	1	336		.59
	Total students			**25,175**	**Weighted average**	**.42**
SAT Verbal, Math, and High School Record	Leonard and Jiang	1995	1	10,000		.49
	Willingham	1985	9	3,442		.53
	Ra	1989	1	170		.58
	Young	1991a	1	1,564		.58
	Ragosta, Braun, and Kaplan	1991	124	2,473	Students without disability	.62
	Total students			**17,649**	**Weighted average**	**.52**

Note: Young (1991a) and Young (1991b) are analyses of the same data, so only the first study is included in Table 1.

between 1980 and the mid-1990s, were studied. Most studies included the entire enrolled class.[3]

The weighted average, reported for all studies of a given predictor or combination of predictors, is the average of the reported correlations weighted by the number of students included in each study.

SAT verbal and mathematical reasoning scores both predict cumulative college grades. The average correlation for high school records is slightly higher than that for the best combination of SAT V and M, a finding that has been observed many times in the past (Ramist, 1984; Wilson, 1983). The combination of the SAT score and high school record has the highest correlation observed. This result is also found commonly in the literature. Note, however, that the weighted average correlation for the best combination of SAT verbal and mathematical measures, .36, is lower than either verbal (.40) or mathematical (.41) alone. This discrepancy occurs because the correlations are based on different samples of students and institutions—if the samples were comparable, the correlation for the combination of V and M would be higher. Like most validity coefficients, the correlations displayed in Table 1 are of moderate size.

A major study of long-term validity did not report comparable correlation coefficients and thus could not be included in Table 1. Bowen and Bok (1998) analyzed data on the academic performance of 32,000 students entering 28 relatively selective undergraduate institutions in 1989. Bowen and Bok's analysis scheme included a set of control variables (race or ethnic group, gender, SES, selectivity of the college, major) as well as SAT scores and high school records in the prediction equation. The correlation with cumulative college rank in class was .45 for the total set of variables. Both SAT scores and high school records contributed significantly to that correlation.

When controlled for gender, race, SES, college selectivity, college major category, and high school rank in class, a 100-point increase in combined SAT V and M scores resulted in a 5.9-point increase in percentile rank in college. This would imply a total of 70 percentile points difference in class rank for students with a com-

bined SAT score in the 400s compared to students with a 1600. Probably because there are few students in the lowest and highest SAT score categories, Bowen and Bok display only the range between <1000 to 1300+, which shows a 20-point difference in class rank between the top and bottom SAT score categories of that range (Figure 3.10, p. 75). This 20-point difference is the largest effect for any variable in the equation. The four next largest effects are the following:

- 16.2 for African American students compared to white students;

- 14.9 for the most selective compared to the least selective of the colleges;

- 14.0 for students whose college major was unknown, compared to humanities majors;

- 10.8 for students in the top 10 percent of their high school class, compared to those in the bottom 90 percent.

In the Bowen and Bok data, the SAT had the highest relationship with cumulative grades and the high school record had the fifth highest relationship. The contributions of SAT and high school record are both statistically significant and substantial.

The Bowen and Bok results were not adjusted for restriction of range, variations in grading standards, or criterion unreliability.[4] Bowen and Bok's results are consistent with the other studies reported here in that the SAT and high school record contribute substantially to the prediction of cumulative college grades. Also, as do the other studies reported here, Bowen and Bok's tends to underestimate the true correlation between predictors and criteria.

Comparison with results for earlier review

Table 2 shows comparative information from Wilson's (1983) review of the literature on predicting the cumulative GPA from preadmission SAT scores and high school records, from more than 40 undergraduate

---

[3]Some studies of student subgroups were included if the subgroups did not appear to differ from the total population. These included students in teacher education programs (American Association of Colleges for Teacher Education, 1992), engineering and computer science majors (Shoemaker, 1986), and students who enrolled for the first time either before age 30 or after age 30 (Moffatt, 1993). Studies of students who may differ from the total enrolled class, such as ethnic or racial minorities or those with disabilities, will be discussed separately below.

[4]The broad categories of college selectivity and college major included in the regression analysis will partially adjust for range restriction and grading stringency. The college selectivity variable will control for range restriction within a set of already selective institutions, but not for the restriction of those institutions compared to all others. The college major variable will control for variations in grading stringency that are related to those particular broad categories, but not for variations that occur at a finer level. There was no adjustment for criterion unreliability.

TABLE 2

Predicting Cumulative Undergraduate GPAs for Classes Graduating Between 1930 and 1980 (Uncorrected Correlations)

Predictor	Paper	Year of Publication	Number of: Institutions	Students	Group, if any	Correlation
SAT Verbal	Wilson	1967	1	259	Women's colleges	.18
	Mauger and Kolmodin	1975	2	838		.35
	Willingham	1962	1	799		.36
	Hardesty	1980	1	1,758		.43
	French	1957	11	4,457		.44
	Wilson	1978; 1980	1	1,200		.48
	French	1957	1	131	Engineering	.49
	Wilson	1978; 1981	1	700		.50
	Total Students			**10,142**	**Weighted average**	**.43**
SAT Math	Wilson	1967	1	259	Women's colleges	.16
	French	1957	1	58	Engineering	.17
	French	1957	11	4,457		.26
	Mauger and Kolmodin	1975	2	838		.31
	Willingham	1962	1	799		.37
	Hardesty	1980	1	1,758		.40
	Wilson	1978; 1981	1	700		.40
	Wilson	1978; 1980	1	1,200		.46
	Total Students			**10,069**	**Weighted average**	**.31**
SAT Verbal and SAT Math	French	1957	1	72	Engineering	.22
	French	1957	1	207	Men	.37
	Olsen and Schrader	1959	3	515	Men's colleges	.42
	Olsen and Schrader	1959	3	681	Women's colleges	.43
	French	1957	1	150	Women	.52
	Total Students			**1,625**	**Weighted average**	**.42**
High School Record	Wilson	1978; 1981	1	700		.34
	Wilson	1967	1	259	Women's colleges	.35
	French	1957	1	153	Women	.39
	Willingham	1962	1	799		.45
	French	1957	11	4,457		.48
	French	1957	1	225	Men	.51
	French	1957	1	124	Engineering	.53
	Wilson	1978; 1980	1	1,200		.56
	Hardesty	1980	1	1,758		.57
	Total Students			**9,675**	**Weighted average**	**.49**
SAT Verbal, Math, and High School Record	Wilson	1976	5	1,905		.41
	Farver, Sedlacek, and Brooks	1975	1	89	White men	.54
				102	White women	.58
	Hills, Bush, and Klock	1964	3	271	Women	.65
			2	267	Men	.67
	Total Students			**2,634**	**Weighted average**	**.47**

Note 1: Table based on Wilson (1983).
Note 2: Wilson (1976): Predictors include Achievement Tests.
Note 3: Farver et al. (1975): Correlation is cross-validated.

institutions and from multiple classes graduating from 1930 through the 1970s. Approximately 12,000 students were included in these studies. The pattern of results for this earlier period is roughly similar to the pattern for more recent studies. One noteworthy difference is in the predictive importance of the SAT mathematical reasoning measure. The average correlation for SAT M is .3 in Wilson, as compared to the average of .4 in more recent studies. An inspection of the earliest studies cited in Wilson (those reported by French, 1957) shows that SAT M correlations are in the .2 range, while all SAT V correlations are in the .4 range. Correlations for SAT M reported from 1962 on were in the .4 range.

This result may reflect the variability to be expected from differences among the institutions and students included in the computations. However, during this time period there was an increase in the number of quantitative and technical courses in the undergraduate curriculum, which increased the importance of mathematical knowledge and reasoning ability for incoming students. An accompanying phenomenon was the increasing number of mathematics courses taken in high school. While college-bound young men have traditionally taken a full load of mathematics in high school, now nearly all high school students, college-bound or not, men and women, minority and nonminority, take the algebra and geometry courses assumed in the SAT mathematical reasoning test. (See, for example, the NAEP report *Trends in Academic Progress* [Campbell, Reese, O'Sullivan, and Dossey, 1996], Table 6.3, p 87.) Both the increased importance of quantitative areas in the college curriculum and the increased level of preparation in high school mathematics for virtually all SAT takers might contribute to the observed increase in validity for the SAT M.

Comparison with results for first-year grades

Another important comparison shows the relationship between the prediction of the first-year GPA and the long-term cumulative GPA. Researchers use the first-year GPA as a measure of success in college for a number of excellent reasons. First-year average is available soon after admission for most of the admitted class. It is often based on a relatively comparable set of required courses, and, as was discussed earlier, grading standards appear to be more comparable in first-year courses than in upper-division courses. The first-year average, how-

ever, does not cover the entire idea of "success in college." Most colleges would like to know whether studies of first-year grades give similar results to more comprehensive criteria such as cumulative grades.

Table 3 summarizes comparisons on two dimensions—studies of cumulative versus first-year GPAs, and studies covering two time periods. The cumulative studies include a review of cumulative GPAs for classes graduating up to 1980 (Wilson, 1983), which is compared to the current review of studies for classes graduating since 1980; the first-year studies include a review of studies by 685 institutions of the first-year GPA for classes entering between the mid-1960s and the early 1980s (Ramist, 1984), which is compared to a recent study of the 1995 entering class at 23 undergraduate institutions enrolling 48,000 students (Bridgeman, Jenkins, and Ervin, 2000). The data in Table 3 are the averages of unadjusted correlations for all institutions included in each review; the correlations are weighted by the number of students included in each study.

The cumulative GPA results are roughly similar in the two time periods, except for the trend for SAT M scores discussed above. As was noted in the discussion of Table 1, the correlation for the best combination of V and M in the more recent studies is lower than the correlations for V and M separately. This anomaly is due to different samples being used in the computations. Comparing studies of cumulative grades to first-year grades, the average correlations over time periods are similar for the best combination of V and M (cumulative $r = .39$; first-year $r = .38$), and for the best combination of SAT scores and high school records (cumulative $r = .49$; first-year $r = .49$). Correlations for individual variables are less comparable but of similar general size. These results provide some confirmation of Wilson's (1983) generalization that studies of cumulative GPAs and first-year GPAs give similar results. Looking at the results for the first-year GPA, it appears that the correlation patterns are also roughly similar for the older and more recent results, except that the correlations are lower in the recent study by Bridgeman et al. (2000).

An exhaustive study of trends in predictive validity by Willingham, Lewis, Morgan, and Ramist (1990) found that a long-term trend of small declines in predictive validity for test scores and high school grades was due to trends unrelated to the test. The researchers found that correlations tend to be lower in less selective institutions (when restriction of range is accounted for), lower when grading standards differ in different courses, and lower when the courses cover a wide range of con-

TABLE 3

Comparing Predictive Validity for Two Criteria in Different Time Periods (Uncorrected Correlations)

| | SAT Scores | | | | SAT + |
	Verbal	Math	V + M	HSR	HSR
Cumulative GPA					
Graduate before 1980: Wilson, 1983	.43	.31	.42	.49	.47
Graduate after 1980: Current review	.40	.41	.36	.42	.52
First-Year GPA					
Enter approx. 1960–1980: Ramist, 1984	.36	.35	.42	.48	.55
Enter 1995: Bridgeman et al., 2000	.30	.30	.35	.36	.44

Note: Data are average institutional correlations weighted by the number of students in the study.

tent from the academic to the practical. An increase in the number of validity studies undertaken by less selective institutions and those with heterogeneous first-year curricula accounted for most of the validity trend.

The research results reported by Willingham et al. (1990) suggest that differences in the size of correlations seen in Table 3 for the earlier and later studies of first-year grades may reflect differences in the two sets of institutions studied. This hypothesis is supported by the following analysis in which correlations were corrected for restriction of range.

Comparison of corrected and uncorrected correlations

Table 4 displays corrected and uncorrected correlations for three multi-institution studies predicting first-year grades and two individual studies predicting cumulative undergraduate grades found in the current review. Bridgeman et al. (2000) studied the classes entering 23 institutions in 1995. Ramist and Weiss (1990) reanalyzed the College Board Validity Study Service archive, focusing on 477 institutions that conducted two or more validity studies between 1970 and 1988. Ramist et al. (1994) studied the entering class of 1985 at 45 institutions. No corrected data were presented in the Wilson (1983) review of long-term validity studies, but results corrected for differing grading standards in cumulative grades were available from two studies (Elliott and Strenta, 1988; Young, 1991a) included in the current review.

Actual correction methods and assumptions differ substantially among these studies. The Young (1991a) and Elliott and Strenta (1988) studies were both concerned with correcting for differences in grading stringency within institutions. They adjusted students'

grades before computing validity coefficients.[5] Since both institutions studied are selective—Stanford and Dartmouth—one would expect the correlations to be artificially low because of the uncorrected effects of restriction of range. Bridgeman et al. (2000) and Ramist and Weiss (1990) corrected for restriction of range in SAT scores and high school records but not for differences in grading standards. Since both studies included institutions with a broad range of first-year course offerings, one would expect the correlations to be artificially low because of the uncorrected effects of differing grading standards. Ramist et al. (1994) corrected for restriction of range in predictors, criterion unreliability, and differences in grading standards. They created a GPA based on corrected grades for the courses actually taken by each student.

Table 4 shows that the studies correcting for either restriction of range or grading standards but not both (Elliott and Strenta, Young, Bridgeman et al., and Ramist and Weiss) reported corrected correlations of about the same size and pattern. The corrected correlation for best combination of SAT scores ranged between .50 and .54 in all three categories; the corrected correlation for SAT scores and high school records ranged between .61 and .64 in all three. Elliott and Strenta and Young, reporting data for two highly selective institutions, did show a somewhat lower corrected correlation for high school records (.41 compared to .55 and .58 in the other two studies). This result is to be expected in such selective institutions, where the range of high school grades or ranks would naturally be low among enrolled students. The absolute size of the corrections in these two studies was relatively small—averaging about +.05. It seems likely that these correlations would be substantially adjusted by a correction for restriction of range. Ramist et al. (1994) showed that the correlations

---

[5]Young did his adjustments within broad categories of disciplines, while Elliott and Strenta based their adjustments on a comparison of students' grades in pairs of courses.

with first-year grades for the most selective third of colleges were higher than those for less selective institutions after range restriction was corrected.

Going back to the question of whether correlations for cumulative grades are similar to correlations for first-year grades, it can be seen that the adjusted correlations for all variables in the first three panels of Table 4 are similar. This provides some support for Wilson's (1983) generalization that studies of first-year grades give similar information to the more difficult long-term validity studies. However, given that only two studies of the cumulative GPA provided adjusted data,

that these two studies were conducted in very selective institutions, and that the adjustments made for the cumulative GPA were different from the adjustments for the first-year GPA, the confirmative effect is limited. More research is needed on this topic.

The Bridgeman et al. (2000) study showed larger corrections than the earlier Ramist and Weiss (1990) study, leading to final correlations of similar sizes for the two time periods. Bridgeman et al. commented that their institutions tended to be rather selective, 7 of the 23 having an average V + M score over 1250. Thus the apparent decline in correlation for the more recent

TABLE 4

Comparing Corrected and Uncorrected Correlations

Predictors	Number of		Weighted Average R:		
	Institutions	Students	Uncorr.	Corr.	Diff.
Cumulative Undergraduate GPA, Classes Graduating Since 1980, Corrected for Differential Grading Standards					
Elliott and Strenta, 1988; Young, 1991a					
SAT, best combination	1	900	.43	.50	+.07
SAT V only	2	2,500	.43	.45	+.02
SAT M only	2	2,500	.42	.50	+.08
High school record (HSR)	2	2,500	.37	.41	+.04
SAT + HSR	1	1,600	.58	.64	+.06
First-Year GPA, Classes Entering 1995, Corrected for Restriction of Range					
Bridgeman et al., 2000	23	48,000			
SAT, best combination			.35	.54	+.19
High school record (HSR)			.36	.55	+.19
SAT + HSR			.44	.61	+.17
First-Year GPA, Classes Entering 1970–1988, Corrected for Restriction of Range					
Ramist and Weiss, 1990	466	At least 600,000*			
SAT, best combination			.37	.52	+.15
SAT V only			.32	.46	+.14
SAT M only			.31	.47	+.16
High school record (HSR)			.48	.58	+.10
SAT + HSR			.54	.64	+.10
First-Year GPA, Classes Entering in 1982 or 1985, Corrected for Restriction of Range, Grading Standards, and Criterion Unreliability					
Ramist, Lewis, and McCamley-Jenkins, 1994	45	48,000			
SAT, best combination			.36	.65	+.29
SAT V only			.30	.60	+.30
SAT M only			.31	.62	+.31
High school record (HSR)			.39	.69	+.30
SAT + HSR			.48	.76	+.28

*Although the number of students was not reported, it could be estimated as follows. Results were reported by size of first-year class (see Ramist and Wiess, 1990, Table 5-10, p. 132). The sample included 205 institutions with first-year classes of fewer than 500 students, 164 with classes between 500 and 1,500, and 95 with classes over 1,500. Conservatively estimating the average class sizes within these ranges as 250, 1,000, and 2,000, the average class size was multiplied by the number of institutions, and then doubled because every institution included had studied at least two enrolled classes.

study shown in Table 3 was probably an artifact based on the greater restriction of range in the institutions studied. No such decline appears in Table 4 when the correlations are corrected for restriction of range.

Finally, the last study reported, Ramist et al. (1994), shows larger corrections and higher correlations than the other studies in the table. Data in this study are corrected for restriction of range, grading standards, and the unreliability of first-year grades. The investigators report an average uncorrected SAT plus HSR correlation with first-year GPA of .48, which becomes .71 when corrected for restriction of range and differential grading standards, and, finally, .76 when corrected also for criterion unreliability.

The evidence for the effect of artifacts on the validity estimates for the cumulative undergraduate GPA is scanty. It is based only on two studies at selective institutions that did not correct either for restriction of range or criterion unreliability. Research will be needed to determine whether more complete corrections, so important to understanding the relationship between the predictors and the first-year GPA, will be equally important to studies of longer-term criteria.

A final summary of the results of the studies of cumulative GPA from the entering classes from the mid 1920s to the early 1990s is presented in Figure 1. Figure 1 combines unadjusted correlations for individual studies reported by Wilson (1983) with the more recent studies reported in this review. Figure 1 shows a box enclosing the middle 50 percent of correlations observed for each predictor. To interpret the box, recall that one quarter of the studies found correlations higher than the top edge of the box, and one quarter of the studies found correlations lower than the bottom edge of the box. The middle 50 percent box not only shows what the correlations are likely to be for typical institutions, they also show the range of results that can be expected. (The individual correlations summarized were weighted by the number of students included in each study.)

Note that uncorrected correlations are used because corrections were not presented in Wilson (1983). Both reviews are based on relatively few studies in differing institutions, and the studies were not corrected for restriction in range, differing grading standards, or criterion unreliability. This leads to undesirable variability in results and probably reduces the size of the observed correlations, but the picture still provides some information. The three single measures— SAT V, SAT M, and HS GPA—all reach about the same level of correlation at the seventy-fifth percentile, while

Typical range of uncorrected correlations*
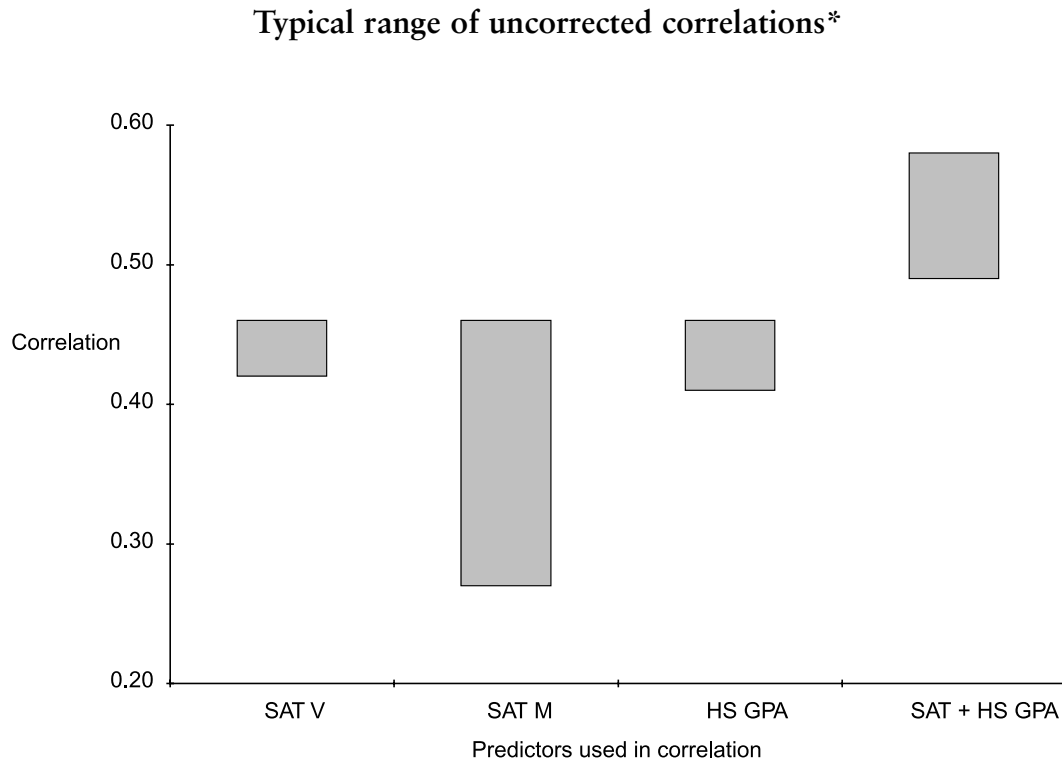
*The middle 50 percent: between the twenty-fifth and seventy-fifth percentile of a size-weighted distribution.

Figure 1. Predicting cumulative undergraduate GPAs.

the combination has noticeably higher correlation at all percentiles. The wide range of correlations observed for the SAT M may be due to variability in the institutions studied, but it may also be due to the changing importance of math in the high school and college curriculum over the last 70 years, discussed above.

Based on studies of first-year grades, one would also expect the correlation of the high school GPA to be somewhat higher than the correlations for the SAT V and M. The somewhat lower than expected correlation for the high school GPA with cumulative college grades may be due to the particular institutions included in this computation. On the other hand, there may be a small substantive difference between predicting first-year college grades from high school grades as compared to predicting cumulative college grades from high school grades. Further research will be needed to answer this question.

Results for subgroups

Women students. Most recent studies of predictive validity for women have concentrated on what is known as "underprediction." When grades are predicted based on a group of men and women, it is commonly found that women's actual grades are slightly higher than predicted while men's grades are slightly lower than predicted. This phenomenon is referred to in the literature as "underprediction" for women and "overprediction" for men. (See, for example, Clark and Grandy, 1984; Cleary, 1992; Friedman, 1989; Hyde, 1981; Hyde, Fennema, and Lamon, 1990; Hyde and Linn, 1986, 1988; Linn and Petersen, 1986; Stanley, Benbow, Brody, Dauber, and Lupowski, 1992; Willingham and Cole, 1997.)

Ramist et al. (1994) found that the average overprediction and underprediction using SAT scores and high school GPAs to predict uncorrected first-year grades was -.06 for women and +.06 for men,[6] meaning that women's predicted grades were .06 grade points lower than their actual grades and men's predicted grades were .06 points higher than their actual grades, based on a 4-point grading scale. For example, a group of women predicted to get a 3.0 first-year grade average would actually average 3.06, while men predicted to receive a 3.0 would actually average 2.94.

Ramist et al. recomputed regression equations within courses to account for differential course-taking patterns for male and women students and differences in grading standards among professors. Within-course

analysis reduced underprediction or overprediction of first-year grades to -.03 for women and +.03 for men. When the selectivity of the undergraduate institution was considered, underprediction for women was shifted somewhat. In selective institutions it was -.01, in average institutions it was -.03, and in less selective institutions it was -.05. Bridgeman et al. (2000) found similar results for first-year grades in the entering class of 1995. Note, however, that Ramist et al.'s within course computations do not adjust for individual student characteristics such as motivation, application, or study habits.[7]

Three studies reviewed presented data on the overprediction or underprediction of the cumulative GPA by gender. They were all conducted at selective institutions (Dartmouth, Stanford, and Berkeley). Table 5 summarizes the results.

Two of the studies also reported correlations corrected for strictness of grading. By correcting, Elliott and Strenta (1988) converted a slight underprediction (-.03) to a slight overprediction (+.02); Young did not report exact corrected figures but stated that he found neither overprediction nor underprediction after correcting. Leonard and Jiang did not report corrected data.

Other studies show inconsistent patterns of grades earned by women in individual courses (Ramist et al., 1994; Strenta, Elliott, Adair, Matier, and Scott, 1994). Strenta et al. (1994) studied science and nonscience course-taking patterns at four highly selective institutions. They found that women earned higher two-year cumulative grades than men in nonscience courses but very slightly lower two-year cumulative grades in science courses. After separating students into those who majored in science in college and those who did not, they found that among nonscience students, women had better science grades than men, while among science students the finding was reversed. The authors comment that

> [t]he resolution of this paradox lies in the difference between standard basic science courses and courses developed by science departments specifically for nonscience students…[which]… carry no major credit and, in most cases, have no labs or prerequisites in math or science (p. 552).

The same pattern of lower science grades for women than for men held also for students who expressed while still in high school an intention to major in science.

African American students. Eight studies analyzed cumulative GPAs for African American students. Most reported significant contributions to the prediction by

---

[6]The signs for overprediction and underprediction as reported in Ramist et al. (1994) have been reversed in this discussion.

[7]Stricker, Rock, and Burton (1993) found that taking into account women's skills as students also reduced the underprediction of their grades.

TABLE 5

Overprediction or Underprediction of Cumulative GPAs for Women and Men (Uncorrected Correlations)

Study	Predictor	Number of: Institutions	Number of: Students	Over (+)/ Under (-) Prediction
Elliott and Strenta, 1988	SAT + HSR	1	927	-.03
Young, 1991a	SAT + HSR	1	1,564	-.08
Leonard and Jiang, 1995	SAT I + SAT II + HSR	1	10,000	-.05

SAT scores and high school records (the American Association of Colleges for Teacher Education [AACTE], 1992; Bowen and Bok, 1998; Friedman and Kay, 1990; Johnson, 1993; Nettles, Thoeny, and Gosman, 1986; Tracey and Sedlacek, 1985). Several investigators reported that the predicted grades for African American students were significantly higher than their actual grades—that is, their grades were "overpredicted" (Bowen and Bok, 1998; Nettles, Thoeny, and Gosman, 1986; Sowa, Thomson, and Bennett, 1989). This phenomenon was discussed at length by Vars and Bowen (1998).

Three studies that compared the correlations for African American and white students are summarized in Table 6.

Unlike a number of major studies predicting first-year grades (Breland, 1979; Bridgeman et al., 2000; Ramist et al., 1994), the correlations in these studies are on average lower for African American than for white students. This may be a true difference between first-year and cumulative grades, but without appropriate adjustments to make the data comparable and with the small and unrepresentative samples presented here, no conclusion is possible.

In the largest study of African American students, Bowen and Bok (1998) analyzed percentile rank in class upon graduation for 28 relatively selective colleges that enrolled 2,300 African American students in the entering class of 1989. These African American students' mean SAT scores (520 verbal and 545 mathematical) were in the top 5 percent of all African American students' SAT scores. Bowen and Bok's regression model included gender, SES, selectivity of college, and major as well as SAT scores and high school records. The differences in criterion and regression model prevented us from including the data in Table 6. The correlation they observed between predictors and college rank in class for African American students in the entering class of 1989 was .44, as compared to the correlation of .45 observed for all students in the class of 1989. As with the full group of students, both SAT scores and high school records contributed significantly to the correlation. The correlation for African American students was equal to the correlation for all students in this 28-institution study. This result is consistent with the Elliott and Strenta study summarized in Table 6, which was also conducted at a selective institution.

As has been found in studies of first-year grades, Bowen and Bok report that the high school record made a smaller contribution to prediction for African American students than for all students combined. This result was also found in the AACTE study in Table 6. In the Bowen and Bok analysis, African American students in the top 10 percent of their high school class had a college rank 5.5 percentage points higher than those in the lower 90 percent. For the total student group, those in the top 10 percent of their high school class had an average college rank 11 percentage points higher. This difference between the African American subgroup and the total group supports other research that finds a

TABLE 6

Comparing Predicted Cumulative GPAs for African American and White Students (Uncorrected Correlations)

Study	Predictor	Number of: Institutions	Number of: AA	Number of: W	Correlations: AA	Correlations: W	Correlations: Diff.
AACTE, 1992	SAT, best combination	23	64	320	.19	.41	-.22
	SAT V				.33	.32	+.01
	SAT M				.08	.26	-.18
	High school record				.33	.48	-.15
Elliott and Strenta, 1988	SAT I + SAT II + HSR	1	46	830	.55	.50	+.05
Tracey and Sedlacek, 1985	SAT, best combination	1	190	1,339	.26	.40	-.16

Note 1: AACTE (1992) sample includes teacher education students only.

TABLE 7

Comparing Predicted Cumulative GPAs for Hispanic and White Students (Uncorrected Correlations)

Study	Predictor	Number of:			Correlations:		
		Institutions	H	W	H	W	Diff.
Pearson, 1993	SAT V	1	200	892	.25	.27	-.02
	SAT M				.30	.30	–

smaller influence of high school records on predicting college grades for African American students (Burton, Morgan, Lewis, and Robertson, 1989; Ramist, 1984).

In contrast to their finding for the high school records, Bowen and Bok found that the SAT was nearly as highly related to cumulative rank in class for African American students as for all students. For all students (holding gender, race, SES, high school rank, college selectivity, and college major constant), the percentile rank in class improved by 20 points from students whose combined SAT scores were less than 1000 to students whose SAT scores were 1300 or more. Within the same SAT range (and holding the same variables constant except for race), African American students' percentile rank increased by 18 points. So the SAT was the strongest predictor for African American students as well as for all students, and the size of the SAT effect was nearly identical.

As was mentioned in the discussion of total group results for Bowen and Bok's study, the correlations reported are likely to be underestimates because of uncorrected effects of restriction of range, differing grading standards, and criterion unreliability. Restriction of range was especially severe for this select group of African American students.

Hispanic students. Pearson (1993) studied four-semester GPAs for Hispanic students (mostly Cuban) at a Florida University. Table 7 summarizes the results. The correlations with the cumulative GPA after two years were essentially the same for both groups. The investigator also found that grades were somewhat underpredicted for Hispanic students when a common prediction equation was used.

Asian American students. Fuertes, Sedlacek, and Liu (1994) studied 431 Asian American students at a large northeastern university. Cumulative grades for semesters 1, 3, 5, and 7 were predicted. The SAT V correlated about .2 with cumulative grades in all semesters; the SAT M correlated between .3 and .4 in all semesters.

Students with disabilities. Ragosta, Braun, and Kaplan (1991) did a major study of students with disabilities, including hearing, learning, physical, and visual disabilities. They gathered data from more than 100 colleges and universities for 4,800 nondisabled students and 1,300 students with disabilities. Table 8 summarizes results of predicting the first-year GPA and cumulative GPA for those students who graduated from college.

Correlations for students with disabilities were generally slightly lower than those for students without disabilities, but they still show a substantial and functionally comparable relationship with cumulative grades. The increase in correlations from first-year grades to cumulative senior year grades for students

TABLE 8

Comparing Predicted First-Year and Cumulative GPAs for Students With and Without Disabilities (Uncorrected Correlations)

Disability	Predictor	Number of:		Correlations:	
		Students	Institutions	First-year GPA	Cumulative GPA
None	SAT, best combination	2,500	100	.51	.52
	SAT + HSR	2,400	100	.59	.62
Hearing	SAT, best combination	50	20	.45	.55
	SAT + HSR	40	20	.61	.66
Learning	SAT, best combination	250	70	.32	.36
	SAT + HSR	200	70	.43	.47
Physical	SAT, best combination	170	30	.45	.44
	SAT + HSR	120	30	.52	.55
Visual	SAT, best combination	100	50	.23	.41
	SAT + HSR	80	50	.28	.50

Table based on Ragosta et al. (1991).

with visual disabilities suggests that they may have some nonacademic difficulties in the first year that are resolved later. A similar pattern, but with smaller differences, can be observed for all students with disabilities. As has been observed for other groups, the combination of SAT scores and high school records is the best predictor for students with disabilities.

Summary for cumulative GPAs

This review tends to confirm the results of Wilson's (1983) earlier review of predictive validity studies for classes graduating between 1930 and 1980. Although both reviews are based on relatively few studies from scattered institutions, average correlations tended to show similar levels and patterns for the commonly used predictors. Both reviews found that SAT scores made a significant and substantial contribution to predicting success in college and that the combination of SAT scores and high school records provided better predictions than either grades or test scores alone.

Only two of the studies of cumulative college grades were adjusted for within-institution grading variations, and none were adjusted for restriction of range or the unreliability of college grades. Unadjusted and adjusted results for studies of first-year GPA suggest that in general the correlations summarized in this review and Wilson's earlier review are likely to underestimate the true relationship between measures used in admission and cumulative college GPAs.

Several investigators studied the long-term validity of the SAT for such groups as women, ethnic or racial minorities, and students with disabilities. In general, the studies found substantial validity in predictions for majority and minority groups. As is found in studies of first-year grades, a pattern of slight underprediction for women and overprediction for African American students was observed. Also as found in studies of first-year grades, the underprediction of women's cumulative college grades was decreased when variations in grading standards were taken into account.

Large studies were done for women, African American students, and students with disabilities. Studies for Asian American and Hispanic students were based on small and possibly unrepresentative samples and should be interpreted with caution. The largest study of African American students (Bowen and Bok, 1998) was based on relatively selective institutions in which the average African American student was in the top 5 percent of all African American SAT test-takers, so even though it was a large study, its results may not generalize to African American students attending more typical institutions.

Very few of the studies of subgroups investigated a combination of test scores and high school records. Subgroups were frequently studied because the investigators were concerned about possible bias in admission for these special groups. But that concern would be better met by studying the entire range of admission measures used rather than focusing solely on the SAT, which is rarely used as the sole determinant for admission decisions. Linn and Werts (1971) demonstrated years ago that prediction studies give misleading results when important measures used in selection are omitted from the prediction study. Furthermore, studies of a *wider* range of predictors than were actually used in admission might allow investigators to identify possible improvements in the admission process.

III. Predicting Graduation

Graduation: Correlational studies

Eight studies correlated admission predictors with four-, five-, or six-year degree attainment in classes graduating between the 1980s and the mid-1990s. Table 9, summarizing these eight studies, in general shows moderate correlations of preadmission measures with eventual graduation. These are lower than the correlations with cumulative GPAs shown in Table 1 and Figure 1. The lower correlations are to be expected since persistence in college and eventual graduation are influenced by nonacademic factors such as finances, motivation, social adjustment, family problems, or

TABLE 9

Recent Studies Predicting Graduation (Uncorrected Correlations)

Predictors	Number of: Institutions	Students	Weighted Average Correlation
SAT, best combination	400	82,000	.33
SAT V only	350	40,000	.27
SAT M only	350	40,000	.27
High school record (HSR)	380	80,000	.29
SAT and HSR	10	15,000	.29

Data based on American Association of Colleges for Teacher Education, 1992, $N = 350$; Astin, 1991, $N = 4,000$; Astin, 1993, $N = 76,000$; Astin, Tsui, and Avalos, 1996, $N = 76,000$; Kanarek, 1989, $N = 12,000$; Tracey and Sedlacek, 1985, 1986, and 1987, $N = 1,200$; Willingham, 1985, $N = 3,400$; Young and Barrett, 1992, $N = 90$. Several of these studies are reanalyses or follow-ups of a single database. Redundant data have been removed from this table.

health. However, these correlations suggest that there is a solid academic component to graduation that is measured by the preadmission measures.

The best combination of SAT scores, studied at the largest number of institutions, was the best predictor of graduation. It was slightly better than the high school record, which was studied at almost all institutions enrolling nearly all the studied students.

Even with the very large numbers studied, there is one anomalous result in Table 9. The multiple correlation for the SAT and high school record with graduation (.29) is lower than the correlation for the best combination of the SAT V and M (.33). This is almost certainly due to differences in the institutions and students used to compute these correlations. Note that only 10 of 400 institutions reported a multiple correlation for the SAT and high school record. In a comparable group of colleges the correlation with graduation for the combination of the SAT and high school record would be slightly *higher* than the correlation for the SAT alone.

Willingham (1985) found that the rates of graduation at different institutions were predictable, although whether or not a given individual would graduate was generally not predictable. In his nine-institution study, a variable identifying which institution the student attended raised the correlation of preadmission measures with graduation from .3 to .4. This result was also observed by Bowen and Bok (1998). Willingham found very low correlations within individual institutions. The average correlation between preadmission predictors and graduation within an institution was .15, which was not significant in six of the nine institutions studied.

These results from Willingham and Bowen and Bok suggest that the correlations observed in Table 9 may be partly due to institutional effects. Most of the results in Table 9 are based on multi-institution studies, so the tendency of more selective institutions to have higher graduation rates will affect the correlations. Pending further research, one cannot be sure what part of a correlation in Table 9 is due to the institution-level relationship of selectivity to retention and what part is due to the predictability of individual students' graduation from their grades and SAT scores.

Kanarek (1989), in her study of a large mid-Atlantic university, added seven statistically significant predictors to the basic test scores plus high school record equation and raised the correlation with graduation from .3 to .6. The largest contributors to the correlation (persistence to the sophomore year and first-year GPA, which accounted for 62 percent of the variance explained by the equation) were data not available at admission. This supports Wilson's (1983) observation that the best predictors are those closest in time and in content to what is being predicted.

Kanarek (1989) also used discriminant function analyses to classify students into graduates and nongraduates. In a model that included both postadmission variables and preadmission variables, Kanarek was able to reach 79 percent correct classifications. The statistically significant postadmission variables were first-year GPAs, persistence to the sophomore year, and basic skills tests in English and math given in the first year. The statistically significant preadmission variables included SAT V and SAT M scores, high school ranks, and a number of self-reported variables from the SAT background questionnaire, including number of years of high school courses in academic subject areas, most recent grades in these subject areas, and self-reported ability in mathematics, writing, and speaking English.

Commenting that postadmission variables are not of help in admission, Kanarek also computed several discriminant function analyses based on preadmission variables alone. The equation based on SAT scores and high school ranks resulted in 64 percent correct classifications overall, significantly lower than the 79 percent possible with postadmission information. However, adding the other preadmission variables listed above only increased correct classifications to 65 percent. When running discriminant function analyses for subgroups (results reported below), Kanarek used the full set of preadmission variables. Different variables separated graduates from nongraduates for different subgroups, but in all cases the most important variables were high school records, SAT scores, or a combination of both.

Three studies attempted to predict interim persistence: return for sophomore year (Kanarek, 1989; Willingham and Breland, 1982), and five-semester persistence (Tracey and Sedlacek, 1987). Correlations were low, ranging from .01 (high school record predicting return for sophomore year) to .17 (SAT M predicting five-semester persistence). Kanarek did not report correlations, but did report an overall correct classification rate of 60 percent for return to sophomore year. Addition of other preadmission predictors raised correlations for interim persistence only slightly.

Bowen and Bok (1998) predicted graduation within six years for nearly 33,000 students attending 28 relatively selective colleges and universities. They used a logistic regression model including gender, ethnic group, SES, selectivity of the college, and whether or not the institution was a women's college, as well as SAT scores and high school records, to predict graduation. The prediction was significant, and the SAT and high school record contributed significantly to it. By making a number of

TABLE 10

Percent Graduating in Four Years, Given Test Scores and High School GPAs

HS GPA	SAT (or Converted ACT) Score Range					
	<700	700–849	850–999	1000–1149	1150–1299	1300+
A, A+	28%	45%	55%	64%	71%	80%
A-	29%	41%	52%	58%	65%	73%
B+	29%	38%	46%	56%	62%	63%
B	21%	32%	39%	46%	51%	48%
B-	17%	26%	33%	35%	44%	38%
C+	17%	18%	24%	27%	28%	—
<C+	10%	16%	19%	21%	—	—

Data based on Astin, Tsui, and Avalos (1996), Table 9, p. 14.

statistical assumptions,[8] we estimate that the correlation between predictors and graduation is between .20 and .24. This estimate is not corrected for restriction of range.

Bowen and Bok also found that students' gender, ethnic group, and SES and the selectivity of the college were significantly related to the probability of graduating within six years. Astin and colleagues published a number of reports based on a major longitudinal follow-up study of the entering class of 1985, including 76,000 students at 365 institutions. Statistical weights were applied to make the results generalizable to all first-time full-time college students in 1985. They reported information in the form of expectancy tables (see the following section) and regression equations (reported in Table 9 and in the series of sections on subgroups, below). To predict graduation for all students and for subgroups, Astin et al. (1996) used complex regression equations with 110 possible variables (including 34 entering-student characteristics; 42 "bridge" variables such as dorm residency and financial arrangements; 10 college type variables; and 24 faculty and peer environment variables).

Graduation: Expectancy tables

Several researchers (Astin, 1991; Astin, Tsui, and Avalos, 1996; and Manski and Wise, 1983) reported expectancy tables showing the probability of graduation as a function of test scores and high school records. The results for Astin et al. (1996) are the most com-plete, based on 76,000 first-year students entering 365 undergraduate institutions in 1985. Their data on 53,000 students for whom there were data on test scores, grades, and graduation status as of spring 1989 are presented in Table 10.

This table shows an important relationship between high school grades, SAT scores, and attainment of an undergraduate degree despite the relatively low correlations of these variables. The percent graduating in four years ranges from 10 percent (both grades and test scores low) to 80 percent (both high).

Similar tables for graduation after six and nine years show that students with low preadmission scores and grades do graduate given more time, and they partially close the gap with students who had better admission credentials. For example, Astin et al. found that the percent graduating in the lowest score and grade category rose from 10 to 21 percent nine years after matriculation. The percent in the highest category also rose, but only by 3 percentage points, from 80 to 83 percent.

Manski and Wise (1983) found that, at a given high school rank, an increase of 1 standard deviation of the sum of SAT V and M scores (about 200 points) was associated with an increase of about 11 percentage points in persistence rate. Similarly, controlling for SAT scores, 1 standard deviation of high school rank also added about 11 percentage points to the persistence rate. Astin (1991) reported that for a given high school GPA level, 150 points on the SAT (combined) increased the rate of four-year degree attainment by about 9 per-

[8]Tucker and Lewis (1973) and Bentler and Bonett (1980) present statistical methods allowing one to estimate a correlation coefficient from chi-squared statistics and their degrees of freedom. These were applied to the logistic regressions reported by Bowen and Bok. It is necessary to have these statistics for a model with an intercept only and a complete model: These allow one to estimate a ratio of explained variance to total variance. Bowen and Bok did not report degrees of freedom for their "restricted" and "unrestricted" models, so we estimated them based on the number of cells that would exist in a completely crossed design using their predictors and criterion, making different assumptions about the number of empty cells to be expected. Whether we assumed that half or three quarters of the 8,640 cells were empty (conservative given the sample size of 32,000) had little effect on the estimated correlation. The correlation we present is the square root of the ratio, which estimates R^2.

centage points. This is equivalent to 12 percentage points for a standard deviation of 200, and it is consistent with Manski and Wise's finding.

Several researchers who reported their results in some form other than correlations or expectancy tables found positive relationships between SAT scores, high school records, and college graduation. They include the American Association of Colleges for Teacher Education (1992), Astin (1991), Bowen and Bok (1998), Crouse and Trusheim (1988), Kanarek (1989), Kane (1994 and 1998), Robinson and Morgan (1989), Willingham (1985), and York, Bollar, and Schoob (1993).

Graduation for subgroups

Women students. In several analyses of the entering class of 1985 (Astin, 1993, and Astin et al., 1996), Astin and colleagues reported that women are slightly more likely to graduate from college than men. Bowen and Bok (1998) also found women more likely to graduate than men. They report (in their sample of 28 relatively selective institutions) that women who attend women's colleges are less likely to graduate within six years. Kanarek (1989) reported that women in a large mid-Atlantic university were slightly more likely to graduate (63 percent in five years) than men (60 percent). Kanarek also found that women's graduation could be predicted slightly more accurately (66 percent overall correct classification) than was possible for men (63 percent correct classifications).

African American students. Studies typically find that African American students have a lower rate of graduation than white students (Astin et al., 1996; Bowen and Bok, 1998; Kanarek, 1989; Kane, 1994).

Bowen and Bok (1998) studied nearly 2,700 African American students who enrolled in 28 relatively selective institutions in 1989. Overall, 75 percent of the African American students graduated from their first college, and 79 percent graduated from some institution within six years, a very high rate compared to other institutions. Bowen and Bok cite National Collegiate Athletic Association (NCAA, 1996) figures of 59 percent graduation overall for the class of 1989 at 305 Division I institutions and 40 percent for African American students. Astin et al. (1996), in their sample weighted to represent all first-time, full-time students entering college in 1985, estimated even lower six-year graduation rates of 47 percent for white students and 31 percent for African American students.

Bowen and Bok's logistic regression on graduation for African American students was significant, and the SAT and high school record contributed significantly to it. Gender, SES, and institutional selectivity also contributed significantly. SES was a more important predictor of graduation for African American students than it was for the combined group of students.

Tracey and Sedlacek (1985, 1986, 1987) found that the SAT was a small but significant predictor of graduation for both African American and white students. Noncognitive questionnaire variables also improved prediction for both African American and white students, but they added more to prediction for African American students. For example, they found approximately equal correlations of about .1 for African American and white students between graduation and the SAT. When their noncognitive questionnaire was added to the SAT, the correlations improved, especially for African American students. The combined correlation was approximately .2 for white students and .4 for African American students. Some of the variables that were predictive for both groups included attitudes or self-appraisals such as "academic self-confidence." Self-report data that were helpful for African American students but not whites included "understanding racism" in the 1985 analysis and other variables in the later analyses. Different factors related to persistence in each study, even though all three studies were based on the same group of students at various points in their undergraduate career.

Kanarek (1989) found that discriminant function analysis correctly classified graduates and nongraduates at about the same rate for African American and white students (approximately 60 percent correct classifications).

Hispanic students. Astin et al. (1996) found that both African American and Mexican American students in the entering class of 1985 had four-year graduation rates about 9 percentage points lower than white students (controlling for SAT/ACT scores, high school records, and gender).

Kanarek (1989) found that a discriminant function analysis of graduates versus nongraduates classified Hispanic, Asian American, and women students slighter better than others: approximately 65 percent overall correct classifications, as compared to 63 percent for men and about 60 percent for white and African American students.

Asian American students. Astin et al. (1996) reported that Asian American students had higher graduation rates than other groups of students. After 9 years, 58 percent of Asian American students had graduated, as compared to 47 percent of white students, 40 percent of Mexican American students, 37 percent of Puerto Rican students, 34 percent of African American students, and 33 percent of Native American students.

Kanarek (1989) found relatively high correct classifications (71 percent) for Asian American graduates

but lower correct classifications (58 percent) for nongraduates, with an overall correct classification rate of 65 percent. Astin et al. also predicted graduation somewhat better for Asian American students (correlation = .48) than for whites (*r* = .43). Fuertes, Sedlacek, and Liu (1994), studying persistence of Asian American students through the fifth and seventh semesters, found SAT M significant at both stages, but SAT V was significant only at the fifth semester.

Native American students. McEvans and Astin (1992) found that Native American students had the lowest graduation rate among all ethnic and racial groups. Astin et al. (1996) reported that 23 percent of Native American students had graduated after four years; 33 percent had graduated after nine years. After controlling for the high school record, SAT/ACT, and gender, Astin et al. showed that Native Americans' graduation rate was 18 percentage points lower than whites' after four years; 13 percentage points lower than whites' after six years; and 10 percentage points lower than whites' after nine years. Astin et al. were able to predict graduation for Native Americans as accurately as they were for whites (*r* = .43 in both cases).

Athletes. Benson (1993) reported that athletes' rate of graduation after five years increased by 8 percentage points (from 48 to 56 percent) after passage of NCAA proposition 48. Benson reported that a 1 standard deviation increase in the high school core course GPA was associated with a 7 percentage-point increase in the graduation rate (holding admission test scores and institutional selectivity constant) and that a 1 standard deviation increase in admission test scores was associated with a 13 percentage point increase in the graduation rate (holding the core GPA and institutional selectivity constant). The study was based on nearly 6,000 NCAA Division I athletes.[9]

Students with disabilities. Ragosta, Braun, and Kaplan (1991), in their study of students with and without disability from over 100 undergraduate institutions, were equally able to predict graduation for disabled and nondisabled students. The results are summarized in Table 11.

Summary for graduation

Preadmission SAT scores and high school records correlate significantly with measures of graduation. Generally moderate correlations were observed for predictions of graduation, lower than the correlations for predictions of cumulative GPA. Several large, representative studies presented expectancy tables rather than correlations, showing a substantial relationship

TABLE 11

Predicting Graduation After Four Years for Students With and Without Disabilities

Predictor	Number of Students	Correct Classifications
Students without Disability		
SAT, best combination	5,092	58%
High school record	4,956	59%
SAT + HSR	4,956	59%
Students with Disabilities		
Hearing		
SAT, best combination	249	64%
High school record	194	66%
SAT + HSR	194	70%
Learning		
SAT, best combination	824	60%
High school record	537	57%
SAT + HSR	535	60%
Physical		
SAT, best combination	497	60%
High school record	323	66%
SAT + HSR	322	67%
Visual		
SAT, best combination	357	57%
High school record	241	61%
SAT + HSR	241	57%

Note: Table based on Ragosta et al., 1991, Table 15, p. 16.

between academic predictors and graduation. For example, a low of 10 percent of students with high school grades below C+ and SAT V plus M below 700 graduated after four years, while 80 percent of students with high school grades of A or A+ and SAT V plus M 1300 or more graduated after four years.

There were only 14 studies of graduation, and several of them were analyses of the same database. However, most of these studies were based on substantial numbers of students and institutions, and a number of them employed representative samples. Four studies analyzed two of a series of longitudinal studies of different high school graduating classes beginning in 1972. These studies, based on scientific samples of students, were sponsored by the National Center for Education Statistics. Several other studies by Astin and colleagues were based on a large study of the 1985 entering class of 365 colleges and universities sponsored by the American Council on Education.

None of the studies of college graduation corrected for restriction of range. A number of the studies employed ordinary least squares regression, which will

[9]Benson employed logistic regression. Institutional selectivity was defined by the mean admission test score (SAT or ACT) at each institution. The core GPA was based on 11 high school courses specified by NCAA.

tend to underestimate the relationship between predictors and the probability that a student will graduate; logistic regression or log-linear analysis is preferable for such data. More studies, correctly adjusting for restriction of range and employing more appropriate regression estimations, will be needed before it can be determined how successful the SAT and high school record are in predicting the probability of graduation.

However, at least two of the studies with large samples, Bowen and Bok (1998) and Willingham (1985), used appropriate regression techniques and studied both cumulative grades and graduation rates for the same students. Both of these studies found lower correlations for graduation than for grades. Bowen and Bok found a correlation between preadmission predictors of .45 for cumulative rank in class, and we estimated a correlation between .20 and .24 for the logistic regression of graduation on preadmission predictors. Willingham, who found a correlation of .53 for SAT scores and high school records when predicting cumulative college GPA, found a correlation of .29 between the same predictors and graduation. While the studies reviewed do not make it possible to estimate the exact size of the correlation for predicting graduation, they do make it clear that the correlation to be expected is lower than the correlation for predicting cumulative college grades.

Lower graduation rates for African American, Hispanic, and Native American students and slightly higher graduation rates for women were reported. Although there were some variations observed in the percentage of graduates among subgroups, the probability of graduation was predicted about as well for all students as for women, African American, Hispanic, Asian American, and Native American students, students with disabilities, and athletes.

One finding that requires further research is the relatively low average correlation ($r = .15$) of admission predictors with graduation found within the individual institutions in Willingham's 1985 study. This result was confirmed in one single-institution study (Tracey and Sedlacek, 1985, $r = .1$, $N = 1,200$) but not in another (Kanarek, 1989, $r = .3$, $N = 12,000$). This effect needs to be studied in institutions with a range of selectivity, because the studies reporting low correlations are for relatively selective institutions.

The persistence criterion appeared to become more predictable as the college career proceeds.

Although this result is based on a very small number of studies, it suggests that many students may leave college in early semesters for nonacademic reasons.

Various studies showed that information about a student's performance in college (first-year grades, persistence to sophomore year) predicted graduation better than the variables available at admission. An institution wanting to improve retention would be well advised to use this information to track student progress. However, admission officers making selection decisions may be happy to learn that preadmission measures do provide dependable information about the probability that students will graduate.

IV. Other Predictors and Criteria of Success

In general, the validity literature reports very few examples of alternative measures of success. Several authors who have considered alternate conceptions of success have commented on how little empirical data is available on important outcomes of higher education (Boyer, 1987; Chickering, 1999; Klitgaard, 1985).[10] These authors have discussed broad conceptions of college outcomes, as have several others (Astin and Panos, 1969; Taber and Hackman, 1976; Whitla, 1981). Further progress in this area of research depends on substantive discussion of goals for admission by the institutions themselves. Research and development can then be organized to provide a broader array of outcome measures that meet institutional needs. A broader array of outcome measures will very likely require new admission measures, especially if the outcomes go beyond academics. This section will discuss both new predictors and new criteria of success, with emphasis on nonacademic measures.

Admission practice has valued a range of academic and nonacademic outcomes (sometimes tacitly) and has traditionally considered a range of academic and nonacademic credentials. Selective colleges are known to consider many factors beside academic qualifications because most of their applicants have high test scores and high school records (Blackburn, 1990). The private liberal arts colleges studied by Willingham (1985) varied in acceptance rate from about 20 to about 90 percent, and they varied in the percentage of applicants

---

[10]There is a larger technical literature evaluating the technical promise of possible measures of success. For example, Hartnett and Willingham (1980) evaluated a number of possible criteria of success in higher education with a view to covering broadly many possible definitions of success in higher education. Campbell, Kuncel, and Oswald (1998) applied theoretical work on the traits underlying complex work performance, developed in industrial and organization psychology since the mid-1980s, to performance in graduate school. They, unlike Hartnett and Willingham, focused on students' traits rather than on institutional values.

with SAT scores over 600 from less than 10 to about 60 percent. All nine of these colleges considered a variety of academic and nonacademic factors in selection.

The methods of considering other factors vary widely from college to college, and the specific factors considered and the subset of factors most valued also vary (Blackburn, 1990). The ratings may be narrative, numerical, or both; the evaluation may be holistic or focused on details; they may be done by admission professionals or faculty or both. In general, however, these evaluations have several elements in common. They are based on a reading of the entire admission file (and thus the effect of any specific student quality is difficult to determine), and they are based on professional judgment. In their study of *Personal Qualities and College Admissions,* Willingham and Breland (1982) found that intended weights on areas of achievement and background were not always consistent with the weights that predicted actual admission decisions.

Willingham (1985) found that ratings by the admission office contributed significantly to predicting various measures of success in college and complemented the more objective ratings done for his research study.

A major challenge is to develop effective connections between admission practices and a broad conception of college outcomes. A solid beginning was made in Willingham's 1985 study entitled *Success in College,* which developed individual institutional definitions of success in collaboration with nine participating undergraduate institutions; developed outcome measures; proposed possible predictors of success; gathered these indicators for an incoming class; monitored its progress through graduation; and evaluated the various proposed predictors. The study included nine private institutions with enrollments between 1,400 and 2,500 with strong liberal arts programs and varying levels of selectivity. It included some 25,000 applicants: about half were admitted, 5,000 enrolled, and 3,500 graduated.

Willingham's Success in College

Willingham's study established that all nine institutions counted scholarship, leadership, and artistic or athletic accomplishment in their definition of success. Analysis of the faculty's final ratings of seniors as most successful overall showed that scholarship, leadership, and accomplishment were about equally valued, and, further, that there was only modest overlap among the three. Of the many preadmission measures studied, only four provided information beyond what is available from the high school record and SAT. These four measures did not substantially improve the prediction of scholarship, but they did improve the prediction of leadership and accomplishment. The four preadmission measures were high school honors, the school reference, the applicant's personal statement, and "follow-through," defined as a student's continuing successful effort in two or more high school activities. In a recent summary of this study, Willingham (1998, p. 14) states that "a major conclusion of the study was that these colleges could not most effectively admit students they regarded as most successful if they selected on school rank and test scores alone."

Table 12 specifies the major predictors and criteria of success studied. In general, the predictors were

Table 12

Major Predictors and Criteria of Success

Predictors of Success
High school rank
SAT scores
High school accomplishments, objective measures: academic honors, community, athletic or creative achievement, follow-through, leadership, and work experience
High school accomplishments, narrative measures: personal statement—writing and content scores; school reference
Students' goals in career, intellectual, creative, physical, leadership and social areas; students' educational aspirations
Admission staff ratings, including outstanding interview, special talents, special attributes, and overall rating
Criteria of Success
Cumulative college GPA (also college GPA, years 1–4)
Graduation (also persistence to senior year, double major, and time to graduation)
Admitted to advanced Ph.D., law, or medical program
Scholarship (either college honors or departmental honors)
Leadership (elected or appointed to office)
Accomplishments (scientific, artistic, physical, organizing, others)
Overall success rating by faculty

Information based on Willingham, 1985.

coded by trained research staff. The criteria of success were defined by institutional committees, and students were nominated in the categories of success by faculty. Student self-ratings and peer ratings were also included.

Table 13 summarizes the major findings of *Success in College*. Note that because the initial correlations are lower, the additional predictors (3–6) have a substantial effect on predicting leadership, accomplishment, and overall success. Measures 3–6 only increase the prediction of scholarship by 7 percent (.04/.57), but they improve the prediction of leadership by 65 percent, accomplishment by 42 percent, and overall success by 25 percent.

In addition to these overall results, Willingham found relationships between specific kinds of leadership and accomplishment in high school and the parallel specific attainments in college. He comments that "[t]hey are a good example of an axiom of learning theory: 'Past behavior best predicts future behavior'" (p. 88).

Finally, Willingham (1985) found that the traditional predictors, the SAT and high school rank, are relatively poor predictors of graduation ($r = .15$) in these nine schools, but better predictors of being accepted to advanced Ph.D., law, or medical programs ($r = .32$). When the student's preadmission report of aspiring to an advanced degree was added, the correlation with acceptance to advanced study rose to .42. Students high on the SAT and high school rank were also more likely to complete double majors and to graduate earlier than others.

Bowen and Bok's
The Shape of the River

Bowen and Bok (1998) analyzed a database called "College and Beyond" which allowed them to report results for such measures of success as attainment of higher degrees, postgraduation income, leadership, and job satisfaction in addition to their analysis of cumulative college rank in class and graduation, reported above. They analyzed academic performance data from the classes of 1976 and 1989 in 28 relatively selective undergraduate institutions. The class of 1989 totaled 32,000 students, 2,300 of whom were African American; the class of 1976 totaled 30,000 students, 1,800 of whom were African American. SAT scores and high school records were significant predictors for all of the criteria studied except post-graduate leadership and job satisfaction. In comparison, Willlingham (1985) studied success within the collegiate years and found the high school record to predict all measures of success and the SAT to predict all measures but college leadership.

In a finding similar to Willingham's, Bowen and Bok found that the SAT was a particularly strong predictor of attaining an advanced degree, even in the presence of measures of college performance, including major field and rank in class. For Bowen and Bok's prediction of attaining a Ph.D. or a law, medicine, or business professional degree, we estimated correlations (using the method outlined in footnote 7) of .39 for all students and .24 for African American students. The SAT was a stronger predictor of graduate degree

TABLE 13

Predicting Four Kinds of Success in College: Which Preadmission Measures Contribute Significantly (X) (Uncorrected Correlations)

Preadmission Measure	Scholarship	Leadership	Accomplishment	Most Successful
1. High school rank	X	X	X	X
2. SAT scores	X		X	X
3. High school honors	X			
4. Follow-through		X	X	X
5. Personal statement	X			
6. School reference		X		X
Correlations				
HSR + SAT	.57	.20	.24	.36
HSR + SAT + measures 3–6	.61	.33	.34	.45
Increase in *R* from adding measures 3–6	.04	.13	.10	.09

Note 1: Table adapted from Willlingham, 1985, Table 5.2, p. 90.

Note 2: Correlations are biserial correlations between actual criterion scores and scores predicted from logistic regression in each college. Results are averaged over the nine colleges studied.

attainment than the high school record.[11] Again confirming Willingham (1985), Bowen and Bok report that another very strong preadmission predictor of attainment of a graduate degree was the student's preadmission aspiration to a higher degree.[12]

The prediction of income was more complicated. In models containing only preadmission measures for both men and women, the SAT, high school rank, SES, and being Asian American all were positively and significantly related to postcollege income ($R = .19$). The SAT was more strongly related to income than was high school rank.[13] The SAT contributed little to the prediction of income for African American men and women, while high school rank in class contributed significantly to prediction of income for African American men but not African American women. Overall, income was somewhat more predictable for African American men ($r = .34$) and women ($r = .24$) than for the total group ($r = .19$ for both men and women) in the model containing only preadmission measures (Bowen and Bok, 1998, Tables D.5.4 and D.5.5, Model 1). In income prediction models including college performance measures and economic sector measures (being employed in the profit sector or self-employed, for example), the SAT and high school record no longer contributed to prediction (see the tables cited above, Models 3, 4, and 5).

Bowen and Bok, like Willingham, found that preadmission academic measures were best at predicting academic success. One nonscholarly success measure, income, has already been discussed: The SAT and high school record contributed significantly to prediction of income only when other more powerful measures were not included, and only for the total group. Other nonscholarly success measures, including job satisfaction and leadership in cultural, community, and youth activities, were virtually unrelated to test scores and high school records. The common predictor variable for these measures was wealth—those with incomes over $149,000 reported being more satisfied and more active. Interestingly, being an African American was an equally important predictor of leadership in cultural, community, and youth activities. African American men and women at all income levels were more likely to participate and take leadership posi-

tions. Whites participated at equal rates in athletic activities only.

In general, the satisfaction and leadership variables were not very predictable even using postcollege information such as income, marriage, and children. Our estimated correlations (see footnote 7) ranged from .01 (leadership in community and social activities) to .18 (job satisfaction).

Summary for other predictors and criteria of success

New predictors. The Bowen and Bok study, based on existing quantitative data at a sample of 28 colleges, did not analyze predictors other than the common test scores and grades. The purpose of their study, to analyze the results of using race in admission, did not require new predictors. It focused on finding a broader definition of success in college. Several studies by Tracey and Sedlacek, evaluating the use of a noncognitive questionnaire, indicated a possibly promising area of research, but their results were not consistent enough to generalize. Willingham's 1985 study in collaboration with nine colleges systematically developed and evaluated a large array of possible predictors of success in college. Two measures, high school honors and the personal statement, made a significant but small contribution to predicting scholarship. Although Willingham judged that this contribution was not large enough to be of importance in admission for the nine colleges studied, it might become important in more selective applications, such as admission to highly selective colleges or scholarship selection.

Two other predictors in the Willingham study— "follow-through" (successful and sustained participation in at least two high school activities) and school references—made significant and substantial contributions to predicting success in areas other than scholarship. The school reference contributed to predicting leadership, and an overall faculty rating of success and follow-through contributed to predicting leadership, accomplishment, and overall success. These are virtually the only proven predictors of important nonscholarly definitions of success in college, and they

[11]This statement is based on odds ratios. Students in the top 10 percent of their high school class had an odds ratio of 1.2 compared to students in the lower 90 percent. In contrast, the odds ratio for students with SAT scores between 1000 and 1099 was 1.6 compared to those with scores under 1000. For every 100 point SAT interval, the odds ratios increased, reaching 3.0 for SAT scores over 1299 (Bowen and Bok, 1998, Table 3.4.2).

[12]The odds ratio was 2.8 for those who indicated they intended to get an advanced degree versus those who did not (Bowen and Bok, 1998, Table 3.4.2).

[13]For all men and all women, the top 300-point intervals reported for the SAT were each worth in the range of $6,000 to $15,000 in income, while being in the top 10 percent of the high school class was worth $6,000 (Bowen and Bok, 1998, Tables D.5.2 and D. 5.3, Model 1).

were shared by nine different institutions. In addition, Willingham showed relationships between specific high school accomplishments and related college accomplishments that could be of importance in selecting students for an institution's areas of specialization.

Attempts to find new measures that predict college success have been difficult. New cognitive measures may correlate very highly with current measures (cf. French, 1957, 1964), although performance measures and evidence of accomplishments may better differentiate among specific skills. New measures may be elaborate, time-consuming, and expensive. Measures of accomplishments have shown promise, but practitioners are concerned about the ease of faking such measures. *Documented* accomplishments (Stricker and Rock, 1996) may provide a means of overcoming fakery problems. Finally, new *predictor* measures are very likely to require new related *outcome* measures since distinctly different predictors are not likely to be highly related to current academic outcome measures.

New criteria of success in college. The relationship between criteria of success in college and institutional values and goals is clear. Of the two studies that evaluated new criteria of success in college, the one (Willingham's) that consulted with institutions developed more predictable measures of success in college. The Bowen and Bok study achieved its stated purpose of exploring the long-term ramifications of using race in selective admissions, but it did not contribute as much to the literature of predictive validity for admission. Of the outcomes that Bowen and Bok explored, the academic criteria of cumulative rank in college class, graduation, and attainment of graduate and professional degrees were the ones best predicted. The prediction of income by admission credentials was significant but small. Better predictions were possible using information about college class rank, major, and economic sector of employment. In the presence of those stronger variables, high school records and test scores no longer contributed to prediction of income. Admission credentials did not predict postcollege job satisfaction or leadership.

In contrast, Willingham, confining himself to measures of success within the college years, developed a number of definitions of success beyond cumulative college record and graduation. These included measures of scholastic honors, leadership, accomplishments in science, art, athletics, organizing, and other areas, an overall rating of success, and admission to doctoral or professional graduate programs. All of these measures of success were predicted by admission credentials, although the prediction of criteria other than scholarship depended largely on new nonacademic predictors.

V. Summary and Discussion

Summary

Predicting cumulative GPAs. This review tends to confirm the results of Wilson's (1983) earlier review of studies predicting cumulative college GPAs. Both reviews are based on relatively few studies from scattered institutions. Both reviews found that SAT scores consistently made a significant contribution to predicting success in college as defined by college grades; both found that the combination of SAT scores and high school records provided better predictions than either predictor alone. This review showed that the information available from studies of first-year grades is generally comparable to the information from studies of cumulative grades, providing some support for Wilson's observation to that effect.

There were studies of substantial size that supported the ability of SAT scores and high school records to predict cumulative college grades for women and African American students and students with disabilities. Smaller studies provided some support for the validity of SAT scores and high school records for African American, Hispanic, and Native American students, and for students over 30 years old.

Predicting graduation. While a relatively small number of studies evaluated the predictability of college graduation, the majority of them were based on large samples of students and colleges and universities, and several covered representative samples of students and institutions. These studies establish that the SAT and high school record are significant predictors of graduation. The correlations observed were moderate and lower than the correlations of admission credentials with cumulative GPA.

Lower graduation rates for African American, Hispanic, and Native American students, and slightly higher graduation rates for women and Asian American students were reported. Although actual graduation rates varied, graduation was as predictable for women, students with disabilities, and athletes, and for African American, Hispanic, Asian American, and Native American students as it was for the total group of students.

Other predictors and criteria of success. For academic measures of success, the traditional test scores and high school records appear to be perfectly adequate predictors. In addition to predicting cumulative grades, the SAT and high school record predicted college or departmental honors, acceptance to graduate or professional school, and completion of a graduate or professional degree. Nonacademic measures of success were

infrequently studied, although it is plausible that the generally lower correlations of SAT scores and high school records with graduation are due to a strong nonacademic component to graduation. Such influences as finances, health, and student personality clearly influence persistence in college. That they influence persistence more than grades seems likely, given the substantial number of college students who withdraw while in good academic standing.

The few studies that addressed nonacademic measures of success showed that the traditional academic predictors, test scores and high school records, have moderate to no relationship to nonacademic success. The only nonacademic success measure after college that was predicted by SAT scores and high school records was income. Within-college nonacademic measures of success included leadership; artistic, athletic, organizational, and civic accomplishments; and an overall faculty rating of success. These nonacademic success measures were significantly predicted by academic predictors, but nonacademic predictors also made a substantial contribution to prediction. Accomplishments in high school, particularly sustained and successful persistence in a few special areas (called "follow-through"), and the school recommendation letter were the strongest predictors of nonacademic success in college.

Discussion

We started this review with a number of questions being asked by admission officers and faculty concerned with undergraduate admission: Are the admission credentials in use good and fair predictors of important outcomes of a college education? Is there current information, based on the populations of students now attending colleges and universities? Are there other potential admission credentials that should be considered? We are now prepared to answer these questions—with yeses, and a few additional comments.

Yes, SAT scores and high school records are good predictors of the academic outcomes of college—cumulative grades, graduating with college or departmental honors, acceptance to graduate school, and attaining a graduate degree. Grades and test scores predict all of these academic accomplishments for a total student body and provide similar levels of prediction for women and African American students. We also have evidence that SAT scores and high school records predict graduation moderately well, although not as well as they predict the other academic variables. For the more frequently studied definitions of success—cumulative grades and attaining an undergraduate degree—we know that the predictions made by SAT scores and high school records provide a similar level of prediction not only for women and African American students but also for students with disabilities. A few studies provide limited evidence of long-term validity for Asian American, Hispanic, and Native American students.

Yes, there is evidence of validity for the kinds of students currently attending college—the current review covers students who graduated from college in the 1980s or 1990s. The studies covered ethnic and racial minority students, women, athletes, and students with disabilities. The studies of the cumulative grade criterion were not as representative as desirable, and there were very few studies for Asian American, Hispanic, and Native American racial and ethnic minorities and older students. The only studies for one very important and growing group of students, non-native English speakers, predicted first-year GPA, but not any long-term criterion of success. In the main, however, there is recent evidence for the importance of SAT scores and high school records in predicting valued outcomes of undergraduate education.

Finally, yes, there are other admission credentials that deserve consideration. There are important nonacademic outcomes of a college or university education, such as leadership, and artistic, athletic, organizational, and civic accomplishments. These nonacademic accomplishments are only partly predicted by high school records and SAT scores—nonacademic credentials contribute substantially to predicting them. The wide variety of talents and performances called for in college suggests that careful consideration should be given to including measures of a broad range of important academic and nonacademic skills and learning styles in the admission process. A broader set of admission measures may also improve prediction for those students who use different skill sets and coping mechanisms to succeed in college. As concerns about fairness grow, new and different predictors of success seem the most likely way to accommodate students with diverse backgrounds, values, and talents. It is not sufficient, however, to add a broader array of *predictors* to the admission decision. The *criteria* of success in college must also be expanded to include these different valued outcomes.

Some caveats about what has been learned. We have learned that the contributions of SAT scores and high school records are statistically significant and of practical utility to admission officers. However, we do not necessarily know just how large the validity coefficients are because the studies reviewed did not apply existing statistical corrections that would base their results on a more comparable set of assumptions. The correlations are reduced by unknown and differing amounts by restriction of range, variations in grading

standards, and by unreliability in the measures of success. (Two of the studies of cumulative grades corrected for variability in grading standards alone.) Since these artifacts vary both between colleges and within colleges, we don't know how much the results will change when they are corrected. We do know that the result of correction will be to increase the correlations. Studies correcting for these artifacts in high school grades and first-year college grades suggest that the effect of correcting cumulative grades might be quite large. Corrected correlations are necessary in practice when institutions want to make comparisons—among groups of students, among teachers, among colleges, or among admission measures or different admission policies.

It is commonly concluded that test scores and grades are the best available predictors of success in college, but they certainly do not account for all the variation in college success. Much of the confusion over the validity of admission tests arises from the disparity in validity results due to different levels of selectivity, heterogeneity in curriculum, variation in grading standards, and unreliability of grading practices. The correlations among SAT scores, high school records, and first-year GPAs, corrected for restriction of range, variations in grading standards, and criterion unreliability can no longer be characterized as "small" or even "moderate." The corrected correlation of .76 that Ramist et al. (1994) found when predicting first-year grades from SAT scores and high school records is *large* (Cohen, 1977).

We do not wish to suggest that grades are so flawed that they should not be used either as predictors or criteria in college admission. The SAT was originally introduced to compensate for some of the problems in high school grades, and in return, grades compensate for some of the limitations of SAT scores. The unreliable and subjective components of grades are compensated by test scores; the narrow focus of test scores is balanced by the wide array of knowledge, skills, attitudes, and interests that go into earning grades. Finally, based on what we have learned about various definitions of success in college, it seems safe to say that a combination of both test scores and grades is a bare minimum set of credentials for predicting the wide array of possible desired outcomes.

The reader may be asking whether it is worthwhile to continue studying longer-term criteria of success in college, given that the first-year GPA is available for more students, available soon after matriculation, and based on more comparable grading standards than grades earned in upper-division courses. The results of this review and Wilson's earlier review suggest that the first-year GPA may be a reasonable surrogate for longer-term criteria in the sense that an institution will

get similar statistical information from either kind of study. We would recommend that this suggestion be confirmed by a larger and more representative study, properly adjusted for statistical artifacts. But in addition, there are good reasons for institutions to study long-term measures of success in college.

If studying longer-term success provokes substantive discussion about institutional goals and definitions of success, it would seem to be well worth doing such a study periodically, simply to re-evaluate institutional goals. This information could also be used to improve current admission measures and encourage the development of new predictors to match newly revealed goals. Furthermore, the practice of validating admission decisions with first-year grades seems to give the widespread impression that admission officers care only about success in the first year and that tests are designed exclusively to predict first-year success. Critics are very quick to use this impression to trivialize the admission process. On the contrary, this review has revealed that SAT scores and high school records predict a range of success measures.

SAT scores and high school records have established their place in the college selection process. This is not to say, however, that institutions that value a breadth of talents and viewpoints and that wish to nurture nonscholarly accomplishments, leadership, and future economic and civic contributions as well as scholarship, can meet their goals relying solely on admission test scores and high school records. Warren Willingham (1998) recently suggested that admission decisions in the broad gray area (among students who are neither clear admits nor clear rejects) should be based principally on the institution's broader goals.

Willingham's advice suggests a view of the college transition process that we will discuss in outline. Institutional decisions to admit are made in the context of a transition process that occurs over several years. Before students apply to a college, fairly extensive selection has already occurred. Many colleges may have to make a relatively small number of decisions to eliminate scholastically overoptimistic applicants in order to define a pool of applicants all of whom are reasonable admits. Beyond the initial pruning, most colleges can safely concentrate on meeting institutional goals beyond adequate academic preparation.

After the institutional offer is extended, the transition process continues. Students decide which offer to accept, choose courses to take, choose a major, choose whether to complete their undergraduate degree, and choose whether to pursue advanced degrees. These decisions are all partially related to high school grades and test scores. The observed predictive validity for admission test scores is also influenced by these decisions.

This view of the transition to college has three major stages:

- The national decentralized phase, controlled by students and their advisers, in which students' known or believed academic and nonacademic characteristics are matched to colleges' known or believed academic and nonacademic characteristics. This stage is very much influenced by test scores (both students' scores and college averages) since they are simple numbers based on a common national scale. College nonacademic characteristics are influential only to the extent that they match students' values and are considered credible.

- The centralized college phase controlled by the admission office. This phase has been studied extensively, although usually in a narrow context that does not reflect the complexity of the decisions actually made. The decisions (not adequately represented in most predictive validity studies) are themselves too often based on narrow information not reflecting the full range of the institution's goals.

- The decentralized within-college phase in which students are sorted and sort themselves into courses and majors. This phase is not well understood.

This view of the transition suggests a revised model for predictive validity studies.

Predictive validity: A revised model for future research. Basically, the above conception of the transition to college suggests two basic evaluation models. One covers the entire transition process and includes only those predictors and criteria that are commonly valued by most institutions and most students: a national and decentralized perspective on validity. The process starts with all potential college-bound students and ends with actual college outcomes for each student. Predictors and criteria will focus on the academic, but nonacademic values may also enter in. For example, the major studies summarized here suggest that some goals may generalize to multiple colleges—such goals as accomplishments and leadership in college and civic and economic contributions after college. It is in the interest of all colleges and universities to evaluate and improve this national system because it has major consequences for that part of the transition process that *is* in the control of the individual institution.

The second evaluation model is more locally focused on decisions within the control of the institution. It may still go beyond the fall and winter weeks in the admission office to include the institution's recruiting, academic advising (including requirements for majors and graduation), and retention practices, since all of these processes affect success in college. It may entail thoughtful debate within the institution about actual definitions of success and how they should be put into practice and research to identify valid predictors and valid criteria for important college goals.

This perspective on the evaluation of success in college recognizes actual practice. A great many validity studies are done by individual institutions with the purpose of evaluating the success of their admission practices. Another major portion of the research literature includes multi-institution studies, literature reviews, and meta-analyses that explicitly or implicitly are concerned about the health of the nation's methods for moving students into and through higher education.

Institutional validity studies that simply compute correlations between test scores and high school records and college grades are more likely to meet national-level needs when properly summarized than they are to meet unique institutional needs. The difficult work to reflect the institution's actual goals and to include the major relevant institutional practices (recruitment, advising, and so on) would improve the usefulness of these studies. At the national level, greater consciousness of how to make data comparable and interpretable would greatly improve outcomes. The existence of national validity studies that can be shown to generalize to a wide variety of colleges and universities will make many of the individual studies now done unnecessary. This implies that national studies should be based on a representative sample of institutions and that the generality of the results to important types of institutions be explicitly tested. At both levels the most important need is to develop and validate new predictors and criteria to capture the broader goals of higher education.

References

American Association of Colleges for Teacher Education (1992). *Academic achievement of White, Black, and Hispanic students in teacher education programs* (Research Report ISSB-0-89333-09409). Washington, D.C: U.S. Government Printing Office.

Astin, A. W. (1991). *Assessment for excellence: Philosophy and practice of assessment*. New York: Macmillan.

Astin, A. W. (1993). *What matters in college? Four critical years revisited*. San Francisco, CA: Jossey-Bass.

Astin, A. W., and Panos, R. J. (1969). *The educational and vocational development of college students*. Washington, DC: American Council on Education.

Astin, A. W., Tsui, L. and Avalos, J. (1996). *Degree attainment rate at American colleges and universities: Effect of race, gender, and institutional type*. Washington, DC: American Council on Education.

Baron, J., and Frank, N. M. (1992). SATs, achievement tests, and high school class rank as predictors of college performance. *Educational and Psychological Measurement 52(4),* 1047–1055.

Benson, M. T. (1993). *A statistical comparison of college graduation of freshman student-athletes before and after proposition 48.* (Research/Technical Report 143). Overland Park, KS: National Collegiate Athletic Association.

Bentler, P. M., and Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structure. *Psychological Bulletin 88(3),* 588–606.

Blackburn, J. A. (1990). *Assessment and evaluation in admission.* New York: College Board.

Bowen, W. G., and Bok, D. (1998). *The shape of the river.* Princeton, NJ: Princeton University Press.

Boyer, E. L. (1987). *College: The undergraduate experience of America.* Scranton, PA: Harper and Row.

Braun, H. I., and Szatrowski, T. H. (1984a). The scale-linkage algorithm: Construction of a universal criterion scale for families of institutions. *Journal of Educational Statistics 9(4),* 311–330.

Braun, H. I., and Szatrowski, T. H. (1984b). Validity studies based on a universal criterion scale. *Journal of Educational Statistics 9(4),* 331–344.

Breland, H. M. (1979). *Population validity and college entrance measures.* (College Board Research Monograph Number 8.) New York: College Board.

Bridgeman, B., Jenkins, L., and 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.

Brogden, H. E. (1946). On the interpretation of the correlation coefficient as a measure of predictive efficiency. *Journal of Educational Psychology 37(2),* 65–67.

Burton, N. W., Morgan, R., Lewis, C., and Robertson, N. (1989). *The predictive validity of SAT and TSWE item types for ethnic and gender groups.* Paper presented at the 1989 annual meeting of the American Educational Research Association and the National Council on Measurement in Education, San Francisco, CA.

Camara, W. (1998). *High school grading policies* (College Board Research Note #4). New York: College Board.

Campbell, J. P., Kuncel, N. R., and Oswald, F. L. (1998). *Predicting performance in graduate school: The criterion problem.* In J. P. Campbell and D. S. Ones, "Selection into I-O Programs: Focus on GRE Validity," symposium presented at the 13th Annual Conference of the Society for Industrial and Organizational Psychology, Dallas, TX.

Campbell, J. R., Reese, C. M., O'Sullivan, C., and Dossey, J. A. (1996). *NAEP 1994 trends in academic progress.* Washington, DC: National Library of Education.

Chickering, A. W. (1999). Personal qualities and human development in higher education: Assessment in the service of educational goals. In S. J. Messick (ed.) *Assessment in Higher Education.* Hillsdale, NJ: Erlbaum, pp. 13–33.

Clark, M. J., and Grandy, J. (1984). S*ex differences in the academic performance of Scholastic Aptitude Test takers* (College Board Report No. 84-8, ETS RR-84-43). New York: College Board.

Cleary, T. A. (1992). Gender differences in aptitude and achievement test scores. In *Sex equity in educational opportunity, achievement, and testing: Proceedings of the 1991 ETS Invitational Conference.* Princeton, NJ: Educational Testing Service, pp. 51–90.

Cohen, A. (1998, April 20). Back to square one. *Time,* pp. 30–31.

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences* (Rev. ed.). New York: Academic Press.

Crews, D. (1993). Processing NTE Core Battery test scores (Internal report). Lock Haven, PA: Lock Haven University of Pennsylvania.

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

Elliott, R., and 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),* 333–347.

Farver, A. S., Sedlacek, W. E., and Brooks, G. C. (1975). Longitudinal predictions of university grades for Blacks and Whites. *Measurement and Evaluation in Guidance 7(4),* 243–250.

French, J. W. (1957). *Validation of the SAT and new item types against four-year academic criteria* (ETS Research Bulletin 57-4). Princeton, NJ: Educational Testing Service.

French, J. W. (1964). New tests for predicting the performance of college students with high-level abilities. *Journal of Educational Psychology, 55(4),* 185–194.

Friedman, L. (1989). Mathematics and the gender gap: A meta-analysis of recent studies on sex differences in mathematical tasks. *Review of Educational Research 59,* 185–213.

Friedman, D. L. and Kay, N. W. (1990). Keeping what we've got: A study of minority retention in engineering. *Engineering Education 80(3),* 407–412.

Fuertes, J. N., Sedlacek, W. E., and Liu, W. M. (1994). Using the SAT and noncognitive variables to predict the grades and retention of Asian American university students. *Measurement and Evaluation in Counseling and Development 27,* 74–84.

Goldman, R. D., and Slaughter, R. E. (1976). Why college grade point average is difficult to predict. *Journal of Educational Psychology 68(1),* 9–14.

Hardesty, L. (1980). *Use of multiple regression to predict academic achievement at a small liberal arts college.* ERIC No. ED 185 960. Reviewed in Wilson, 1983.

Hartnett, R. T., and Willingham, W. W. (1980). The criterion problem: What measure of success in graduate education? *Applied Psychological Measurement 4(3),* 281–291.

Hillegas, M. B. (1912) A scale for the measurement of quality in English composition by young people. *Teacher's College Record 13,* 331–384.

Hills, J. R., Bush, M. L., and Klock, J. A. (1964). Predicting grades beyond the freshman year. *College Board Review 54,* 23-25. New York: College Board. Reviewed in Wilson, 1983.

Hyde, J. S. (1981). How large are cognitive gender differences? A meta-analysis using w^2 and d. *American Psychologist 36(8),* 892–901.

Hyde, J. S., Fennema, E., and Lamon, S. J. (1990). Gender differences in mathematics performance: A meta-analysis. *Psychological Bulletin 107,* 139–155.

Hyde, J. S., and Linn, M. C. (Eds.) (1986). *The psychology of gender: Advances through meta-analysis.* Baltimore, MD: Johns Hopkins University Press.

Johnson, R. E. (1993). *Factors in the academic success of African American college males.* Unpublished Doctoral Dissertation. Clemson University.

Kanarek, E. A. (1989). *Exploring the murky world of admissions predictions.* Paper presented at the Annual Forum of the Association for Institutional Research, Baltimore, MD.

Kane, T. J. (1994). *Race, college attendance and college completion* (Research Report R 117- E). Washington DC: Office of Educational Research and Improvement.

Kane, T. J. (1998). Racial and ethnic preferences in college admissions. In C. Jencks and M. Phillips, eds., *The Black-White test score gap.* Washington, DC: Brookings Institution Press, pp. 431–456.

Klitgaard, R. (1985). *Choosing elites.* New York: Basic Books.

Leonard, D. K., and Jiang, J. (1995). *Gender bias in the college predictions of the SAT.* Paper presented at the Annual Meeting of American Educational Research Association and the National Council for Measurement in Education, San Francisco, CA.

Linn, M. C., and Petersen, A. C. (1986). A meta-analysis of gender differences in spatial ability: Implications for mathematics and science achievement. In J. Hyde and M. Linn (Eds.), *The psychology of gender: Advances through meta-analysis.* Baltimore, MD: Johns Hopkins University Press, pp. 67–101.

Linn, R. L. (1966). Grade adjustments for prediction of academic performance: A review. *Journal of Educational Measurement 3(4),* 313–329.

Linn, R. L., and Werts, C. E. (1971). Considerations for studies of test bias. *Journal of Educational Measurement 8,* 1–4.

Manski, C. F., and Wise, D. A. (1983). *College choice in America.* Cambridge, MA: Harvard University Press.

Mauger, P. A. and Kolmondin, C. A. (1975). Long-term predictive validity of the Scholastic Aptitude Test. *Journal of Educational Psychology,* 25, 66–69. (Reviewed in Wilson, 1983.)

McEvans, A., and Astin, A. W. (1992). *Minority student retention rates: Comparative national data from the 1984 freshman class.* Los Angeles: Higher Education Research Institute, University of California at Los Angeles.

Moffatt, G. K. (1993). *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 Center for Education Statistics (NCES) (1984). High school course grade standards. *NCES Bulletin* (Report No. NCES-221b). ED 252 570. Washington DC: National Center for Education Statistics.

National Collegiate Athletic Association (1996). *NCAA Division I graduation-rates report.* Shawnee Mission KS: National Collegiate Athletic Association.

Nettles, M. T., Thoeny, A. R., and Gosman, E. J. (1986). Comparative and predictive analyses of Black and White students' college achievement and experiences. *Journal of Higher Education 57(3),* 289–318.

Nickens, Herbert W. (April, 1998). Questions and answers on affirmative action in medical education. Washington DC: American Association of Medical Colleges.

Office of Educational Research and Improvement (OERI) (1994). *What do student grades mean? Differences across schools.* ERIC # ED 367 666. Washington DC: Office of Educational Research and Improvement.

Olsen, M., and Schrader, W. B. (1959). T*he use of preliminary and final Scholastic Aptitude Test scores in predicting college grades* (SR-59-19). Princeton, NJ: Educational Testing Service.

Pearson, B. Z. (1993). Predictive validity of the Scholastic Aptitude Test (SAT) for Hispanic bilingual students. *Hispanic Journal of Behavior Sciences 15(3),* 342–356.

Ra, J. B. (1989). Validity of a new evaluative scale to aid admissions decisions. *Evaluation and Program Planning 12,* 195–204.

Ragosta, M., Braun, H., and Kaplan, B. (1991). *Performance and persistence: A validity study of the SAT for students with disabilities* (College Board Report 91-3). New York: College Board.

Ramist, L. (1984). Validity of the ATP Tests. In T. Donlon (Ed.), *The College Board technical handbook for the Scholastic Aptitude Test and Achievement Tests.* New York: College Board, pp. 141–170.

Ramist, L., Lewis, C., and McCamley, L. (1990). Implications of using freshman GPA as the criterion for the predictive validity of the SAT. In W. W. Willingham, C. Lewis, R. Morgan, and L. Ramist (Eds.), *Predicting college grades: An analysis of trends over two decades.* Princeton, NJ: Educational Testing Service, pp. 253–288.

Ramist, L., Lewis, C., and 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.

Ramist, L., and Weiss, G. (1990). The predictive validity of the SAT, 1964 to 1988. In W. Willingham, C. Lewis, R. Morgan, and L. Ramist (Eds.), *Predicting college grades: An analysis of institutional trends over two decades.* Princeton, NJ: Educational Testing Service, pp. 117–140.

Robinson, G. E., and Craver, J. M. (1989). *Assessing and grading student achievement.* Arlington, VA: Educational Research Service.

Robinson, P. W., and Morgan, J. A. (1989). *A study of the relationship between an entering freshman's Scholastic Aptitude Test scores and her persistence to graduation at Brenau* (Unpublished Practicum Papers). Fort Lauderdale, FL: Nova University.

Roser, M. A. (December 28, 1998). 10% college admission law yields mixed results. *Austin-American Statesman.*

Roser, M. A. (September 15, 1998). Hispanic, Black freshmen gain ground at UT. *Austin-American Statesman.*

Roser, M. A. (January 19, 1998). UT paints two pictures of minority enrollment. *Austin-American Statesman.*

Shoemaker, J. S. (1986). *Predicting cumulative and major GPA of UCI engineering and computer science majors.* Paper presented at the Annual Meeting of the American Educational Research Association and the National Council on Measurement in Education, San Francisco CA.

Sowa, C., Thomson, M. M., and Bennett, C. T. (1989). Prediction and improvement of academic performance for high-risk Black college students. *Journal of Multicultural Counseling and Development 17(1),* 14–22.

Stanley, J. C., Benbow, C. P., Brody, L. E., Dauber, S., and Lupowski, A. E. (1992). Gender differences on eighty-six nationally standardized aptitude and achievement tests. In N. Colangelo, S. Assouline, and D. Ambroson (Eds.) *Talent development: Proceedings from the 1991 Henry B. and Jocelyn Wallace National Research Symposium on Talent Development.* Unionville, NY: Trillium Press, pp. 41–48.

Strenta, A. C., and Elliott, R. (1987). Differential grading standards revisited. *Journal of Educational Measurement 24(4),* 281–291.

Strenta, A. C., Elliott, R., Adair, R., Matier, M., and Scott, J. (1994). Choosing and leaving science in highly selective institutions. *Research in Higher Education 35(5),* 513–547.

Stricker, L. J., and Rock, D. A. (1996). *Measuring accomplishments: Pseudoipsativity, quantity vs. quality, and dimensionality* (ETS RR-96-8). Princeton, NJ: Educational Testing Service.

Stricker, L. J., Rock, D. A., and Burton, N. W. (1993). Sex differences in predictions of college grades from Scholastic Aptitude Test scores. *Journal of Educational Psychology 85(4),* 710–718.

Taber, T. D., and Hackman, J. D. (1976). Dimensions of undergraduate college performance. *Journal of Applied Psychology 61(5),* 546–558.

Tracey, T. J., and Sedlacek, W. E. (1985). The relationship of noncognitive variables to academic success: A longitudinal comparison by race. *Journal of College Student Personnel 26(5),* 405–410.

Tracey, T. J., and Sedlacek, W. E. (1986). *Prediction of college graduation using noncognitive variables by race.* Paper presented at the Annual Meeting of the American Educational Research Association and the National Council on Measurement in Education, San Francisco CA.

Tracey, T. J., and Sedlacek, W. K. (1987). *A comparison of White and Black student academic success using noncognitive variables: A Lisrel analysis* (Research Report 6-87). College Park, MD: University of Maryland.

Tucker, L. R. (1960). *Formal models for a central prediction system* (ETS RB-60-14). Princeton, NJ: Educational Testing Service.

Tucker, L. R., and Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika 38(1),* 1–10.

Vars, F. E., and Bowen, W. G. (1998). Scholastic Aptitude Test scores, race, and academic performance in selective colleges and universities. In C. Jencks and M. Phillips (Eds.) *The Black-White test score gap.* Washington, DC: Brookings Institution Press, pp. 457–479.

Whitla, D. K. (1981). *Value added and other matters.* ERIC No. ED 228 245. Cambridge, MA: Harvard University, Office of Instructional Research and Evaluation.

Wightman, L. F. (1997). The threat to diversity in legal education: An empirical analysis of the consequences of abandoning race as a factor in law school admission decisions. *New York University Law Review 72(1),* 1–53.

Willingham, W. W. (1962). *Longitudinal analysis of academic performance* (Research Memorandum 62-5). Georgia Institute of Technology. Reviewed in Wilson, 1983.

Willingham, W. W. (1974). Predicting success in graduate education. *Science 183,* 273–278.

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

Willingham, W. W. (1998). *Validity in college selection: Context and evidence.* Paper presented at a workshop on the role of tests in higher education admissions sponsored by the National Research Council, Washington, DC.

Willingham, W. W., and Breland, H. M. (1982). *Personal qualities and college admissions.* New York: College Board.

Willingham, W. W., and Cole, N. S. (1997). *Gender and fair assessment.* Hillsdale, NJ: Erlbaum.

Willingham, W. W., Lewis, C., Morgan, R., and Ramist, L. (Eds.) (1990). *Predicting college grades: An analysis of institutional trends over two decades.* Princeton NJ: Educational Testing Service.

Willingham, W. W., Pollack, J., and Lewis, C. (2000). *Grades and test scores: Accounting for observed differences.* (ETS RR-00-15). Princeton, NJ: Educational Testing Service.

Wilson, K. M. (1967). *Type of analysis underway or planned: A selective review.* College Research Center, Vassar College. Cited in Wilson (1983).

Wilson, K. M. (1976). The utility of a standard composite for forecasting academic performance in several liberal arts colleges. *Research in Higher Education 5,* 192–213.

Wilson, K. M. (1978). *Predicting the long-term performance in college of minority and nonminority students: A comparative analysis in two collegiate settings* (ETS RB 78-6). Princeton, NJ: Educational Testing Service.

Wilson, K. M. (1980). The performance of minority students beyond the freshman year: Testing a "late bloomer" hypothesis in one state university setting. *Research in Higher Education 13,* 23–47.

Wilson, K. M. (1981). Analyzing the long-term performance of minority and nonminority students: A tale of two studies. *Research in Higher Education 15,* 351–375.

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

Wolfe, R. N., and Johnson, S. D. (1995). Personality as a predictor of college performance. *Educational and Psychological Measurement 55(2),* 177–185.

York, M. C., Bollar, S., and Schoob, C. (1993). *Causes of college retention: A systems perspective.* Paper presented at the Annual Meeting of the American Psychological Association, Toronto, Canada.

Young, J. W. (1990). Adjusting the cumulative GPA using item response theory. *Journal of Educational Measurement 27(2),* 175–186.

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(3),* 229–239.

Young, J. W., and Barrett, C. A. (1992). Analyzing high school transcripts to improve prediction of college performance. *Journal of College Admission 137,* 25–29.

Ziomek, R. L., and Svec, J. C. (1997). High school grades and achievement: Evidence of grade inflation. *NASSP Bulletin 81(95-3),* 105–113.