Applied Measurement in Education

Routledge
Taylor & Francis Group

ISSN: 0895-7347 (Print) 1532-4818 (Online) Journal homepage: http://www.tandfonline.com/loi/hame20

Beyond Correlations: Usefulness of High School GPA and Test Scores in Making College Admissions Decisions

Richard Sawyer

To cite this article: Richard Sawyer (2013) Beyond Correlations: Usefulness of High School GPA and Test Scores in Making College Admissions Decisions, Applied Measurement in Education, 26:2, 89-112, DOI: 10.1080/08957347.2013.765433

To link to this article: http://dx.doi.org/10.1080/08957347.2013.765433

Published online: 09 Apr 2013.

Applied Measurement in Education, 26: 89–112, 2013

ISSN: 0895-7347 print/1532-4818 online
DOI: 10.1080/08957347.2013.765433

Beyond Correlations: Usefulness of High School GPA and Test Scores in Making College Admissions Decisions

Richard Sawyer
ACT, Iowa City Iowa

Correlational evidence suggests that high school GPA is better than admission test scores in predicting first-year college GPA, although test scores have incremental predictive validity. The usefulness of a selection variable in making admission decisions depends in part on its predictive validity, but also on institutions' selectivity and definition of success. Analyses of data from 192 institutions suggest that high school GPA is more useful than admission test scores in situations involving low selectivity in admissions and minimal to average academic performance in college. In contrast, test scores are more useful than high school GPA in situations involving high selectivity and high academic performance. In nearly all contexts, test scores have incremental usefulness beyond high school GPA. Moreover, high school GPA by test score interactions are important in predicting academic success.

Citing correlational evidence, many people believe that high school grades are better than scores on college admission tests in predicting first-year college GPA, but test scores have incremental predictive validity. For example, Kobrin, Patterson, Shaw, Mattern, and Barbuti (2008) reported correlations of .36 for high school GPA, .35 for SAT scores, and .46 for high school GPA and SAT scores jointly. Evidence for ACT scores (1999, 2008b) is similar: For the academic years 1970–1971 through 2006–2007, typical multiple correlations of high school subject-area grade averages with first-year college GPA ranged from .48 to .51. The high school grade average correlations were .01 lower to .08 higher than the corresponding ACT score correlations, and were .04 to .09 lower than the corresponding correlations for ACT scores and high school grades jointly. Of course, there is complexity behind the general result; for example, SAT scores have higher correlations with first-year GPA at selective and small institutions (Kobrin et al., 2008).

From this evidence, one might conclude that high school grades are more useful than test scores in making admission decisions, but that test scores have incremental usefulness. This conclusion, however, is oversimplified. As will be discussed below, the usefulness of a selection variable for admission to college does depend in large part on its predictive power, but it also depends on admission officers' goals. Correlations are related to the variance in an outcome variable that is explained by predictor variables. Admission officers, however, are not typically interested in explaining variance: They are interested in achieving their institutions' larger goals to educate students successfully. Usefulness also depends on other statistical issues,

Correspondence should be addressed to Richard Sawyer, Ph.D., ACT, 500 ACT Drive, P.O. Box 168, Iowa City, IA 52243. E-mail: richard.sawyer@act.org

such as utility, applicant self-selection, and institution selectivity. In this article, I show that the issue of usefulness is more complex and more interesting than is implied by correlations or other regression statistics. I show that in many cases, the conventional wisdom based on correlations does apply to usefulness, but that in some important respects, it does not.

This article proposes certain types of statistical evidence related to usefulness in making admission decisions. Such evidence would be relevant in supporting some of the assumptions in a validity argument (Kane, 2006). Of course, a comprehensive validity argument would need to address many other considerations, such as the student characteristics required for academic success in college, the constructs purported to be measured by high school grades and test scores, the extent to which these variables actually do measure their purported constructs, and the extent to which they measure irrelevant constructs.

ADMISSION GOALS

To gauge the usefulness of a selection variable in achieving a goal, we need to specify the goal. Institutions require test scores in admission for several reasons. Obviously, some reasons relate to promoting academic success: enrolling students who are most likely to succeed, making course placement decisions, and making scholarship decisions Another reason is objectivity: to include in decision making a component that can be interpreted the same way for all applicants (thus, serving as a balance to the lack of standardization in high school grades).[1] Other reasons include shaping a brand image, abetting recruitment, supporting counseling and guidance services, and acquiring data for institutional self-study (Breland, Maxey, Gernand, Cumming, & Trapani, 2002; Lenore-Jenkins & Goff, 2008).

Moreover, institutions' admission decisions also relate to considerations other than academic achievement (Camara, 2005; National Association for College Admission Counseling, 2006; Perfetto, 1999): These considerations include assembling a student body with non-academic achievements, promoting cultural diversity, and promoting support from alumni. Therefore, test scores, high school course work, high school grades, and other measures of academic achievement are only part of institutions' admission decision making. Most institutions make admission decisions using a holistic approach, rather than by an explicit formula (Breland et al., 2002).

Two common goals related to academic achievement are: (1) To maximize academic success among enrolled students and (2) To identify accurately those applicants who could benefit from attending the institution, and to enroll as many of them as possible.

These goals seem similar, but they are not identical. As will be described below, the first goal is related to the proportion of applicants who would succeed academically if they enrolled (success rate). The second goal is related to the proportion of applicants whom an institution correctly identifies as likely to succeed or likely to fail (accuracy rate). Both goals, however, pertain only to institutions with some degree of academic selectivity in their admission policies, rather than to institutions with open admission policies.

[1] In principle, an institution could statistically adjust high school GPA for high school effects (such as in a hierarchical model), thereby making it more standardized. The practical benefits of doing this are limited, however, if admission test scores are also used (Willingham, 2005).

The admission selection strategy for accomplishing the first goal is relatively straightforward: So long as there is a positive relationship between the selection variable and the success criterion, an institution can increase the success rate of its enrolled students by admitting only applicants with the highest values of the selection variable. Highly selective institutions can easily pursue this goal because they typically have many more academically qualified applicants than they can admit. Less academically qualified students tend not to apply to these institutions to begin with; these potential applicants in effect select themselves out of the applicant pools. As a result, variance in the predictors is restricted at these institutions, typically resulting in smaller regression slopes and predictive correlations than at less selective institutions (see, e.g., Kobrin & Patterson, 2010).

Most, if not all, institutions would ideally like to maximize academic success among their enrolled students. Publicly supported institutions, regional institutions, and other institutions that are moderately selective also consider the second goal to be important, however. These institutions' mission is to educate a broad portion of the population, not just the most academically able. A principal goal in their admission decision making, therefore, is to distinguish between applicants who are likely to be successful from those who are not likely to be successful. This goal is more difficult to achieve than the first, because it depends not only on maximizing academic success among enrolled students, but also on making sure that applicants who are denied admission would have had little chance of succeeding if they had enrolled. As will be shown later, it is possible that a selection variable can achieve the first goal (maximizing academic success), but not the second goal (accurately identifying potentially successful applicants).

ACADEMIC SUCCESS

In this article, academic success is defined jointly by retention through the first year and by overall first-year college GPA (*FYGPA*). Other researchers (e.g., Bowen, Chingos, & McPherson, 2009; Saupe & Curs, 2008) have studied long-term academic success (degree completion and cumulative GPA). Long-term success is clearly an important goal for all institutions. Attaining this goal through admission selection, however, is likely to be more feasible at highly selective institutions that attract and enroll only the most academically qualified applicants than at institutions whose mission is to educate a broad segment of students with diverse academic skills. At highly selective institutions, graduation rates and cumulative GPAs are typically very high, and many more highly qualified students apply than can be admitted. At less selective institutions, grades during the first year strongly mediate predictions of long-term success based on pre-enrollment measures (Allen, Robbins, Casillas, & Oh, 2008). To achieve their goals, these institutions select applicants who have a reasonable chance of succeeding in the first year, given the interventions (e.g., counseling and course placement) that the institutions might provide.

INDICATORS OF USEFULNESS AND TARGET POPULATION

Previous work on the usefulness of a test score in making decisions has emphasized calculating the expected value of a cost, benefit, or utility function with respect to the joint distribution of the test score and an outcome variable. For example, Cronbach and Gleser (1965) adapted linear regression models to estimate the expected costs and benefits of using a test score for classifying or selecting individuals. Petersen and Novick (1976) proposed computing an expected "threshold" (0/1) utility function in a Bayesian decision model. Calculating expected costs, benefits, or utilities provides a theoretically firm way to evaluate the usefulness of various decision rules, but the downside is that one must first unambiguously specify a cost, benefit, or utility function. Instead of attempting to specify precisely the utility of outcomes, one could instead calculate simple statistical indicators that are plausibly related to typical goals (Sawyer, 2007). This approach frames the discussion in terms that users readily understand, rather than on an abstract utility function.

Admission selection rules are applied to applicants. Therefore, a common-sense indicator of the usefulness of particular selection rules is the estimated proportion of applicants for whom the admission goals would be achieved if they enrolled (Sawyer, 2007). For institutions whose goal is to maximize academic success among enrolled students by selecting the applicants who are most likely to be successful (Goal 1), this proportion is the estimated success rate:

$$SR\,(c) = \frac{\sum\limits_{i \geq cN} \hat{p}_{(i)}}{(1\ -\ c)N} \tag{1}$$

$$= \text{estimated proportion of applicants who, if enrolled, would be successful.}$$

Here, $\hat{p}_{(1)} \leq \hat{p}_{(i)} \leq \hat{p}_{(N)}$ are the ordered estimated conditional probabilities of success of the N applicants, given one or more selection variables, and $1 - c$ is equal to the proportion of applicants selected. I refer to the variable c as the "cutoff proportion"; it is equal to the cumulative relative frequency associated with a value of the selection variable. The variable c relates to an institution's selectivity in admission; the variable $SR(c)$ is the estimated success rate among the students the institution does admit.

An institution serving a high-risk population might be concerned about minimizing the proportion of academic failure and near-failure (Fs and Ds) among its first-year students; it would define success as a 2.0 or higher GPA. An institution with few academic failures might instead be interested in maximizing the proportion of students who earn a 3.0 or higher average. A highly selective institution that expects most of its students to attain excellent academic achievement might consider higher levels of GPA.

For institutions whose goal is accurately to identify potentially successful applicants (Goal 2), the relevant indicator is the estimated accuracy rate:

$$AR\,(c) = \frac{\sum\limits_{i<cN} [1 - \hat{p}_{(i)}] + \sum\limits_{i \geq cN} \hat{p}_{(i)}}{N} \tag{2}$$

$$= \text{proportion of applicants for whom a correct admission decision is made.}$$

The AR indicator corresponds to an expected utility in which admitting an applicant who would be successful and denying admission to an applicant who would be unsuccessful are both weighted 1, and admitting an applicant who would be unsuccessful and denying admission to an applicant who would be successful are both weighted 0. The AR indicator can be generalized to other utilities by assigning different weights to the four outcomes (Sawyer, 1996). For example, if an institution believes that admitting students who do not succeed is less serious an error than denying admission to students who would have been successful, then it could assign a larger weight to the former outcome than to the latter.

Note that both indicators pertain to the target population of *applicants*, as a whole or in part, rather than only to enrolled students. The indicator *AR(c)* pertains to the entire applicant population for an institution, whereas *SR(c)* pertains to the subset of applicants who meet the institution's cutoff proportion. Predictive validity studies that only summarize correlations and other regression statistics based on data of enrolled students overlook this point. On the other hand, even though the indicators in Equations 1 and 2 pertain to applicants, we must estimate their conditional probability of success component from the data of enrolled students; the reason is that we can obtain outcome data only from applicants who enroll at an institution.

Institutions are unlikely to forego using high school GPA and test scores in making admission decisions solely to do research.[2] We therefore need to make an additional assumption about the conditional probability of success component in the indicators. We assume that the conditional probability of success, given the selection variable(s) X, is the same for the non-enrolled applicants as for the enrolled students:

$$p(x) = \mathrm{P}\left[S = 1 | X = x, \text{non-enrolled}\right] = \mathrm{P}\left[S = 1 | X = x, \text{enrolled}\right],$$

where $S = 1$ if a student is unsuccessful, and $S = 0$ otherwise. This assumption is analogous to that for the traditional adjustment of correlations for restriction of range, which requires that the applicant and enrolled student groups have the same conditional mean and variance functions (e.g., Lord & Novick, 1968). Although empirically testing the correctness of this assumption is not feasible, one could investigate the robustness of the indicators of usefulness to specified departures from the assumption.

With this assumption, calculating the indicators is straightforward: Simply score the entire applicant population with the fitted conditional probability of success function (e.g., using a logistic regression model), and then calculate the indicators from the ordered estimated conditional probabilities.

Another assumption is that this study is anchored in the current environment of admission, in which applicants to most four-year institutions must provide scores on admission tests. A broader analysis of the usefulness of test scores in admission would need to consider what would occur if test scores were not used at all. It is very likely that without test scores, high school grades would be subject to inflationary pressure, thereby eroding their predictive power and, ultimately, their usefulness in achieving institutions' goals. This article does not attempt to model the effects of inflation in high school grades in a hypothetical environment either without college admission tests or with tests optional. There is considerable evidence, however, that even in the current environment, high school grades are subject to inflation (Geisinger, 2009; Woodruff & Ziomek, 2004).

[2]Although some non-open-enrollment institutions have test-score-optional admission policies, most of them continue to use test scores, at least for some applicants (Breland et al., 2002). Furthermore, it is doubtful that any institution disregards high school GPA (or its transformation to high school rank) in its admission decisions.

CUTOFF PROPORTION

Note that the indicators considered here assume that an institution admits the top proportion *1-c* of the applicants, based on their conditional probability of success, and that it denies admission to the bottom proportion *c*. Unlike the correlation coefficient, which is a global measure, the indicators of usefulness described here depend explicitly on *c*. These indicators, as well as other considerations (such as capacity for the number of enrolled students), can inform an institution's choice of *c*.

Of course, institutions do not use simple cutoffs in making admission decisions. Equations 1 and 2 are mathematical idealizations that enable us to compare the properties of alternative selection variables. In this article, cutoff proportions range from .01 (virtually all applicants admitted) to .99 (extreme selectivity).

CORRELATION COEFFICIENT

According to Equations 1 and 2, the indicators *SR(c)* and *AR(c)* are functions of the conditional probability of success *p(x)*, the distribution of *p(x)* in the applicant population, and the cutoff proportion *c*. If the selection variable *X* and the underlying outcome variable *Y* have a bivariate normal distribution in the applicant population, then we can more directly calculate *SR(c)* and *AR(c)* by integrating the bivariate normal density function over appropriate regions of *X* and *Y*:

$$\begin{aligned} SR(c) &= \frac{\int_{z_S}^{\infty}\int_{z_c}^{\infty}\phi(x,y;\rho)dx\,dy}{1-c} \\ AR(c) &= \int_{-\infty}^{z_S}\int_{-\infty}^{z_c}\phi(x,y;\rho)dx\,dy + \int_{z_S}^{\infty}\int_{z_c}^{\infty}\phi(x,y;\rho)dx\,dy, \end{aligned} \tag{3}$$

where z_c is the value of a standard normal variable corresponding to the cutoff proportion c on the selection variable X, $z_S = (y_S - \mu_Y)/\sigma_Y$ is the value of a standard normal variable corresponding to the success level y_S on the outcome variable Y, and ϕ is the bivariate standard normal density function with correlation parameter ρ.

A possible reason why researchers often interpret ρ as a measure of usefulness in selection is that given a bivariate normal distribution, *SR* and *AR* depend on ρ. As is clear from Equation 3, however, *SR* and *AR* also depend on the success level y_S and the cutoff proportion c, as well as on ρ. Moreover, as we shall soon see, the assumption of bivariate normality is not tenable in the context of college admission selection. In particular, high school GPA has a severe negative skew and a pronounced ceiling.

INCREMENTAL USEFULNESS

A basic question that we should ask when evaluating the usefulness of a selection variable *X* is: Does using a variable *X* for selection increase the success rate and the accuracy rate over what would result if an institution did not use X (i.e., if it admitted all applicants or denied admission to all applicants)? Admitting all applicants amounts to setting $c = 0$ in Equations 1 and 2; the resulting quantity for either indicator is the base success rate

$$BSR = SR(0) = AR(0) = \frac{\sum\limits_{all\ i} \hat{p}_{(i)}}{N}, \tag{4}$$

which is also equal to the overall (marginal) probability of success. Denying admission to all applicants amounts to setting $c = 1$, and would result in an accuracy rate of *1-BSR*. I refer to admitting all applicants and denying admission to all applicants as "null decisions."

Note that *SR* depends on both the conditional probability of success function p and on the distribution of p (or, alternatively, of X) in the applicant population. Nevertheless, if p is an increasing function of x (the value of the selection variable X), then *SR* is also an increasing function of x (see Appendix), and therefore of c. Hence, if p is an increasing function of X, *SR(c)* exceeds *BSR*. Incremental usefulness in terms of success rate therefore corresponds to the traditional notion of positive regression slope or correlation.

As we shall see, however, the same need not be true of *AR*: Even though a variable is positively related to success, using it in selection might result in a decrease in classification accuracy. One can show by a straightforward differentiation argument (see Appendix) that *AR(c)* exceeds both *BSR* and *1-BSR* at some cutoff proportion if, and only if, the conditional probability of success function *p(x)* crosses 0.5. Moreover, *AR(c)* achieves its maximum value at the cutoff proportion c' associated with x' (and this maximum exceeds *BSR* and *1-BSR*) if, and only if, $p(x') = 0.5$.

Another important question involves comparing alternative selection variables X and W: Does using X and W jointly increase the success rate and accuracy rate over that which would occur if we used X only? This question pertains to incremental usefulness, and is analogous to the traditional notion of incremental predictive validity.

Because *SR(c)* and *AR(c)* depend on the cutoff proportion c, the answers to both questions can vary, depending on c. A selection variable might have incremental usefulness (either with respect to null decisions or with respect to another selection variable) at some cutoff proportions, but not at others.

DATA

The analyses in this article are based on data from 192 four-year postsecondary institutions that use ACT scores in their admission procedures (ACT, 2008a). The institutions provided outcome data either through their participation in ACT's predictive validity service or through participation in special research projects. The outcome data pertain to the following entering freshman class years: 2003 (1% of institutions), 2004 (31%), 2005 (68%), and 2006 (1%). For institutions that had data from more than one entering freshman class, I used the most recent data.

When students register to take the ACT, they report their high school course work and grades. The analyses are based on *HSAvg*, the average of students' self-reported grades in standard college-preparatory courses, and on *ACT-C*, the Composite (average) of students' ACT scores in English, mathematics, reading, and science. The first-year GPA data were reported by the postsecondary institutions.

Institutional Characteristics

The institutions in the sample are broadly representative of four-year institutions in the United States with respect to their proportion of minority students (.21) and students' average ACT Composite scores (21.8). Fifty-seven percent of the institutions in this study are public, however, whereas only about 30% of all postsecondary institutions are public. Moreover, the institutions in this study tend to be much larger (median undergraduate enrollment 2,883) than four-year institutions generally (median undergraduate enrollment 1,545).

Correlations. The median correlations of *FYGPA* with *HSAvg* and *ACT-C* show the typical result: *HSAvg* is a better predictor of *FYGPA* (median correlation .48) than *ACT-C* (median correlation .44), but *ACT-C* has incremental predictive validity (median multiple correlation .54). Of more interest is the huge variation among institutions in their correlations. At two institutions, *HSAvg* and *ACT-C* were both negatively correlated with *FYGPA*. At neither institution were the correlations statistically significant ($p < .05$), even though the sample sizes were moderately large ($n = 145$ and 154). At the other extreme, *HSAvg* and *ACT-C* jointly accounted for nearly two-thirds of the variance in *FYGPA* (multiple $R = .83$ and .84) at two other institutions. Correlations for both predictor variables were higher at private institutions, at institutions with smaller percentages of minority students, and at institutions with larger standard deviations in the predictor variables. Correlations were lower at highly selective institutions.

Student Characteristics

Table 1 provides summary information about *HSAvg, ACT-C,* and *FYGPA* for the enrolled students in the sample. The mean values of *HSAvg* (3.42) and *ACT-C* (22.6) are slightly higher than those of ACT-tested enrolled students nationally (3.36 and 21.8, respectively). The mean value of *FYGPA* (2.78) is slightly lower than that of ACT-tested enrolled students nationally.

As was noted previously, *SR* and *AR* are functions of the correlation ρ (and of other properties of the selection and outcome variables) when they have a bivariate normal distribution. Figure 1 shows histograms of *HSAvg, ACT-C,* and *FYGPA* standardized to z-scores with respect to the enrolled student population pooled over institutions. Note that the *FYGPA* levels of 2.0, 3.0, 3.5, 3.7 lie in a band from approximately 1 *SD* below the *FYGPA* mean to 1 *SD* above the *FYGPA* mean. Figure 1 also shows a reference curve for the standard normal distribution.

As is clear from the marginal distributions shown in Figure 1, the assumption of bivariate normality is untenable. The distribution of *HSAvg* in our data has a pronounced negative skew (-0.9). The modal category of *HSAvg* is its maximum category, 0.75 to 1.25 standard deviations above the mean. *HSAvg* is also negatively skewed in broader populations: (-0.7) combined

TABLE 1
Distribution of Enrolled Student Characteristics (120,338 Students; 192 Institutions)

Variable	*Mean*	*SD*
HSAvg	3.42	0.50
ACT-C	22.6	4.30
FYGPA	2.78	0.95

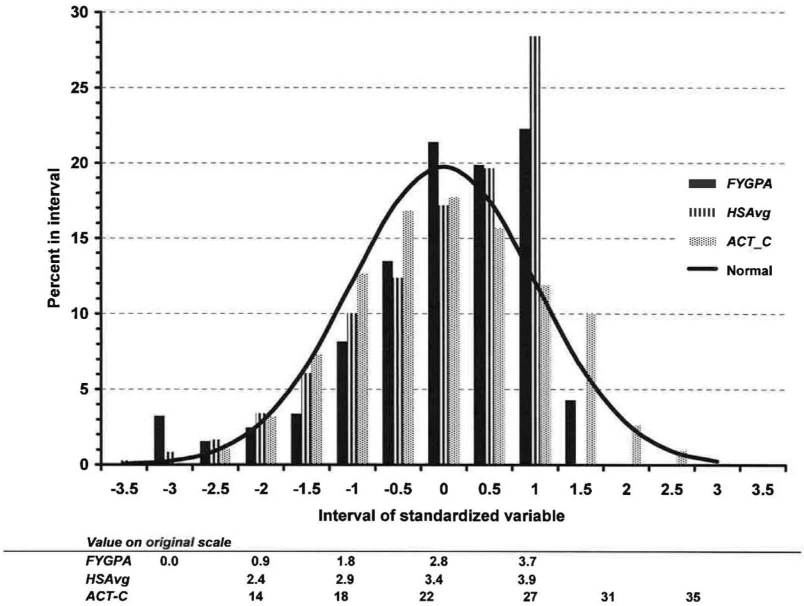

FIGURE 1 Distributions of standardized *FYGPA, HSAvg*, and *ACT-C*, compared to the standard normal distribution.

score sender/enrolled student population; (−0.7) all 2005 ACT-tested high school graduates; and (−0.4) a nationally representative sample of eleventh-grade students (Casillas et al., 2010).

FYGPA is also negatively skewed (−1.1), and has a minor mode at its minimum category (2.75 to 3.25 standard deviations below the mean). The distribution of *ACT-C* is more nearly symmetric (skewness = 0.1). Furthermore, both the conditional mean of *FYGPA*, given *HSAvg*, and the conditional mean of *FYGPA*, given *ACT-C*, are slightly curvilinear (not shown in Figure 1). Therefore, one should be cautious in applying Equation 3 to calculate *SR* and *AR*.

Applicants and Score Senders

As was previously noted, postsecondary institutions make admission decisions about applicants; therefore, indicators of usefulness should be calculated for this target population. The total applicant population consists of non-enrolled applicants (those who are either not admitted or who choose not to enroll) and of enrolled applicants. Data were available for the enrolled applicants, but for several reasons it is not feasible in a study involving many institutions to identify their non-enrolled applicants. I instead used non-enrolled score senders (students who sent their ACT scores to particular institutions, and who did not later enroll) as a proxy for the non-enrolled applicants. The 192 institutions in the sample for this study had 483,451 non-enrolled score senders, in addition to their 120,338 enrolled students.

Applicants could be identified at a subsample of 53 institutions. Analysis of data from these institutions indicated that the score senders had lower *HSAvg* and *ACT-C* means than the applicants did. This result suggests self-selection by students with respect to *HSAvg* and *ACT-C*.

METHOD

Outcome Variables

For the analyses in this article, students who complete the first year with a given level or higher of *FYGPA* are considered to be successful ($S = 1$); otherwise, they are considered to be unsuccessful ($S = 0$). Although dichotomizing a quasi-interval variable such as *FYGPA* degrades information in a statistical sense, dichotomies correspond more closely to admission officers' interpretations: Are a college student's grades high enough at least to get by, or has the student performed well? Moreover, as will become apparent, the usefulness of high school GPA and test scores depends strongly on which level of success one considers.

The analyses in this article consider four levels of success:

- *S20*: Retention through first year, and 2.0 or higher *FYGPA* (minimal success)
- *S30:* Retention through first year, and 3.0 or higher *FYGPA* (typical level of success)
- *S35:* Retention through first year, and 3.5 or higher *FYGPA* (high level of success)
- *S37:* Retention through first year, and 3.7 or higher *FYGPA* (very high level of success)

By these criteria, students who either drop out or have a low *FYGPA* during their first year are unsuccessful. In the data on which this study is based, about 84% of students were at least minimally successful, about 52% were at least typically successful, about 27% were highly successful, and about 16% were very highly successful.

Modeling Probability of Success

According to Equations 1 and 2, an essential component of the indicators *SR(c)* and *AR(c)* is $p(x) = \hat{P}[S = 1 \mid X = x]$, the estimated conditional probability of success, given the value of the selection variable X. I estimated the conditional probability of success using a hierarchical logistic regression model:

$$\ln\left(\frac{p_{ij}(x)}{1 - p_{ij}(x)}\right) = \beta_{0j} + \beta_{1j}x_{ij} \tag{5}$$

The symbol x_{ij} refers to a selection variable (either *HSAvg* or *ACT-C*) for student i at institution j, and the regression coefficients β_{0j} and β_{1j} vary randomly among institutions (see discussion below). To facilitate interpretation of the intercept β_{0j}, as well as computation, I centered x_{ij} about its mean across students (grand-mean centering).

I also estimated models based on *HSAvg* and *ACT-C* jointly. A standard way to model the relationship between probability of success and both variables is to include them as main effects x_{1ij} and x_{2ij}. It is possible, however, that the relationship between the log-odds and each selection variable depends on the value of the other selection variable; we can test this possibility with an interaction term $x_{1ij}x_{2ij}$ in the following model:

$$\ln\left(\frac{p_{ij}(x)}{1 - p_{ij}(x)}\right) = \beta_{0j} + \beta_{1j}x_{1ij} + \beta_{2j}x_{2ij} + \beta_{3j}x_{1ij}x_{2ij} \tag{6}$$

For example, Equation 6 says that the slope of the log-odds on x_{1ij}, namely $\beta_{1j} + \beta_{3j}x_{2ij}$, depends on x_{2ij}.[3]

I estimated both models hierarchically, in which the regression coefficients vary across institutions both systematically and randomly. In Model 6, for example,

$$\begin{aligned}
\beta_{0j} &= \gamma_{00} + \sum_s \gamma_{0s} W_{0sj} + u_{0j} \\
\beta_{1j} &= \gamma_{10} + \sum_s \gamma_{1s} W_{1sj} + u_{1j} \\
\beta_{2j} &= \gamma_{20} + \sum_s \gamma_{2s} W_{2sj} + u_{2j} \\
\beta_{3j} &= \gamma_{30} + \sum_s \gamma_{3s} W_{3sj} + u_{3j}
\end{aligned} \tag{7}$$

For example, the intercept β_{0j} for institution j is the sum of a fixed effect γ_{00} that is constant across institutions, the weighted sum of various institutional characteristics W_{0sj} associated with institution j, and a random effect u_{0j} that is specific to institution j.

The potential level-2 predictors $W_{0sj}, W_{1sj}, W_{2sj}, W_{3sj}$ for $\beta_{0j}, \beta_{1j}, \beta_{2j}, \beta_{3j}$, respectively, were the institution means $\bar{x}_{1j}$ and $\bar{x}_{2j}$ of the centered selection variables x_{1ij} and x_{2ij}, the institutional mean of the centered interaction term, and the institutional characteristics affiliation (public/private), undergraduate enrollment, and percentage minority. I estimated models for which all the student-level fixed effects $\gamma_{00}, \gamma_{10}, \gamma_{20}, \gamma_{30}$ were statistically significant ($p < .001$), the fixed effects $\gamma_{0s}, \gamma_{1s}, \gamma_{2s}, \gamma_{3s}$ for the institution characteristics $W_{0sj}, W_{1sj}, W_{2sj}, W_{3sj}$ were statistically significant ($p < .01$), and the variances of the random effects $u_{0j}, u_{1j}, u_{2j}, u_{3j}$ were statistically significant ($p < .01$).

Initially, I estimated the interaction Model 6 using data of all 120,338 enrolled students in the sample. Plots of the fixed effects in this initial model revealed that for values of *HSAvg* below 2.0, the estimated linear predictor was a weakly decreasing function of *ACT-C*. I therefore re-estimated the interaction model using data only for students with *HSAvg* above 2.0. For the 945 students whose *HSAvg* was less than 2.0, I set the probability of success equal to their overall base success rate.

I estimated Models 5, 6, and 7 with the HLM software (Raudenbush, Bryk, Cheong, & Congdon, 2004), but calculated the indicators of usefulness *SR(c)* and *AR(c)* in SAS. To simplify the data processing, I re-estimated the hierarchical models with PROC NLMIXED (SAS Institute, 2008), which can score the data. NLMIXED requires gargantuan computer resources for large data sets, however, and I could not use it to estimate a model with a random effect u_{3j} in the interaction term coefficient β_{3j}. To calculate success rates and accuracy rates, I used a simplified model without a random effect for the interaction term.

[3]I thank Professor Joseph Rodgers, Vanderbilt University, for his suggestion to estimate interaction models.

Indicators of Usefulness

From the estimated probabilities of success returned by NLMIXED for Models 5, 6, and 7, I calculated *SR(c)* and *AR(c)* using the cutoff proportions $c = .01, .10, .20, .30, .40, .50, .60, .70, .80, .85, .90, .95$, and $.99$ for each selection variable. From the calculated values of *SR(c)* and *AR(c)*, I calculated the following statistics related to the 192 institutions in the data:

Median incremental success rate with respect to base rate:

$$Med.\,Inc.\,SR(c) = \underset{1\le i\le 192}{median}\left\{SR_i(c) - BSR_i\right\}. \tag{8}$$

Median incremental success rate of ACT-C with respect to HSAvg:

$$Med.\,Inc.\,SR^{[ACT-C]}(c) = \underset{1\le \mathrm{i}\le 192}{median}\left\{SR_i^{[HSAvg\ \mathrm{and}\ ACT-C]}(\mathrm{c}) - \mathrm{SR}_\mathrm{i}^{[\mathrm{HSAvg}]}(\mathrm{c})\right\} \tag{9}$$

Frequency of incremental accuracy with respect to null decisions, globally and at cutoff c:

$$Rel.\,Freq.\,Inc.\,Acc. = \text{proportion of institutions whose probability-of-success curve crosses 0.5 somewhere} \tag{10}$$

$$Rel.\,Freq.\,Inc.\,Acc(c) = \text{proportion of institutions for which } AR_i(c) - max[BSR_i, 1 - BSR_i] > 0 \tag{11}$$

Median incremental accuracy rate among the institutions where it is positive:

$$Med.\,Inc.\,AR(c) = \underset{i:AR_i(c)-\max[BSR_i,1-BSR_i]>0}{median}\left\{AR_i(c) - \max[BSR_i, 1 - BSR_i]\right\}, \tag{12}$$

Median incremental accuracy rate of ACT-C with respect to HSAvg:

$$Med.\,Inc.\,AR^{[ACT-C]}(c) = \underset{i:AR_i{}^{[HSAvg\ \mathrm{and}\ ACT-C]}(\mathrm{c})-\max[\mathrm{BSR_i},1-\mathrm{BSR_i}]>0}{median}\left\{AR_i^{[HSAvg\ \mathrm{and}\ ACT-C]}(c) - AR_i^{[HSAvg]}(c)\right\} \tag{13}$$

RESULTS

Hierarchical Models

Table 2 summarizes the hierarchical models for each of the four success levels. Note that in both of the single-variable models (labeled A and B in Table 2), the fixed effects for the *HSAvg* and *ACT-C* slope coefficients are positive and statistically significant ($p < .001$). Moreover, the slope coefficients for *HSAvg* and *ACT-C* both increase with success level. For example, the *HSAvg* slope

TABLE 2
Hierarchical Models for Predicting Success From *HSAvg* and *ACT-C*

			Success Level							
			2.0		*3.0*		*3.5*		*3.7*	
Model	*Level 1 Variable*	*Level 2 Variable*	*Coeff.*	*p <*	*Coeff.*	*p <*	*Coeff.*	*p <*	*Coeff.*	*p <*
Fixed Effects										
A	*Intercept*	*Intercept*	1.983	.001	−0.002	.971	−1.584	.001	−2.539	.001
		Mn_HSAvg	1.734	.001	. . .	. . .	. . .	. . .	−0.783	.004
	HSAvg	*Intercept*	1.582	.001	2.277	.001	3.043	.001	3.72	.001
		Mn_HSAvg	0.009	.001	0.010	.001	0.013	.001	0.019	.001
B	*Intercept*	*Intercept*	1.957	.001	0.127	.005	−1.252	.001	−2.015	.001
		Mn_ACT-C	0.158	.001	. . .	. . .	. . .	. . .	−0.076	.001
	ACT-C	*Intercept*	0.158	.001	0.231	.001	0.272	.001	0.3	.001
		Mn_ACT-C	. . .	. . .	. . .	. . .	−0.008	.001	−0.009	.002
C	*Intercept*	*Intercept*	2.205	.001	0.21	.001	−1.436	.001	−2.409	.001
		Mn_ACT-C	0.112	.001	. . .	. . .	. . .	. . .	−0.136	.001
	HSAvg	*Intercept*	1.535	.001	1.934	.001	2.389	.001	2.864	.001
	ACT-C	*Intercept*	0.105	.001	0.152	.001	0.169	.001	0.185	.001
		Mn_ACT-C	. . .	. . .	. . .	. . .	−0.008	.001	−0.007	.003
	HSAvg X ACT-C	*Intercept*	0.085	.001	0.098	.001	0.089	.001	0.081	.001

		Success Level							
		2.0		*3.0*		*3.5*		*3.7*	
Model	*Level 2 Residual*	*Std. dev.*	*p <*	*Std. dev.*	*p <*	*Std. dev.*	*p <*	*Std. dev.*	*p <*
Random Effects									
A	*Intercept*	0.740	.001	0.635	.001	0.542	.001	0.650	.001
	HSAvg	0.258	.001	0.573	.001	0.795	.001	1.059	.001
B	*Intercept*	0.691	.001	0.574	.001	0.501	.001	0.537	.001
	ACT-C	0.050	.001	0.056	.001	0.056	.001	0.063	.001
C	*Intercept*	0.739	.001	0.639	.001	0.573	.001	0.637	.001
	HSAvg	0.244	.001	0.518	.001	0.71	.001	0.918	.001
	ACT-C	0.048	.001	0.039	.001	0.038	.001	0.053	.001
	HSAvg X *ACT-C*	. . .	. . .	0.044	.001	0.058	.001	0.071	.002

coefficient for the 2.0 success level is 1.582; for the 3.7 level, it is 3.720. This result suggests that *HSAvg* and *ACT-C* are more strongly related to high levels of success than they are to low levels of success.

In the single-variable models, the variances of the *HSAvg* and *ACT-C* slope coefficients among institutions (lower half of Table 2) also increase with success level. This result indicates that there is more variability among institutions in these variables' relationships with higher levels of success than with lower levels of success.

A typical way to compare predictor variables is to standardize their slope coefficients with respect to their standard deviations. On multiplying the fixed effects for the *HSAvg* and *ACT-C* slopes in Table 2 by the standard deviations of the associated predictors, we find that the

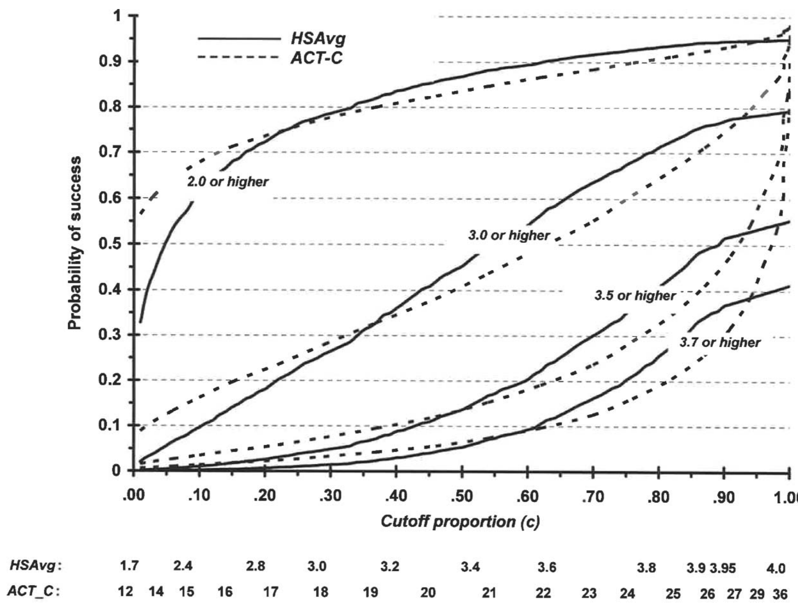

FIGURE 2 Probability of success, given *HSAvg* or *ACT-C*.

standardized regression coefficients for *HSAvg* are uniformly larger than those for *ACT-C*. As was previously noted, however, the usefulness of selection variables depends on other properties, in addition to their regression coefficients. Examining the entire probability of success curves, rather than just the slope coefficients, lets us observe differences in ability to differentiate likelihood of success across the entire ranges of predictor variables.

Figure 2 shows probabilities of success calculated from the fixed effects of *HSAvg* and *ACT-C* in the single-predictor models (5). These probability curves pertain to typical postsecondary institutions (i.e., those for which the random effects are 0). The horizontal axis is scaled in terms of the values of the selection variables and their associated cutoff proportions (cumulative relative frequencies, *c*).

To describe and compare the statistical relationships shown in Figure 2, I noted the following characteristics:

- What range of estimated probabilities is associated with the entire range of the selection variable?
- Does the probability curve cross 0.5?
- Over what values of the selection variable is the probability curve steepest?

By these criteria *HSAvg* is more predictive than *ACT-C* for the 2.0 or higher success criterion. On the other hand, *ACT-C* is somewhat more predictive than *HSAvg* for the 3.0 success level at high cutoff proportions, and much more predictive than *HSAvg* for the 3.5 and 3.7 success levels. This result is consistent with that reported by Noble and Sawyer (2004).

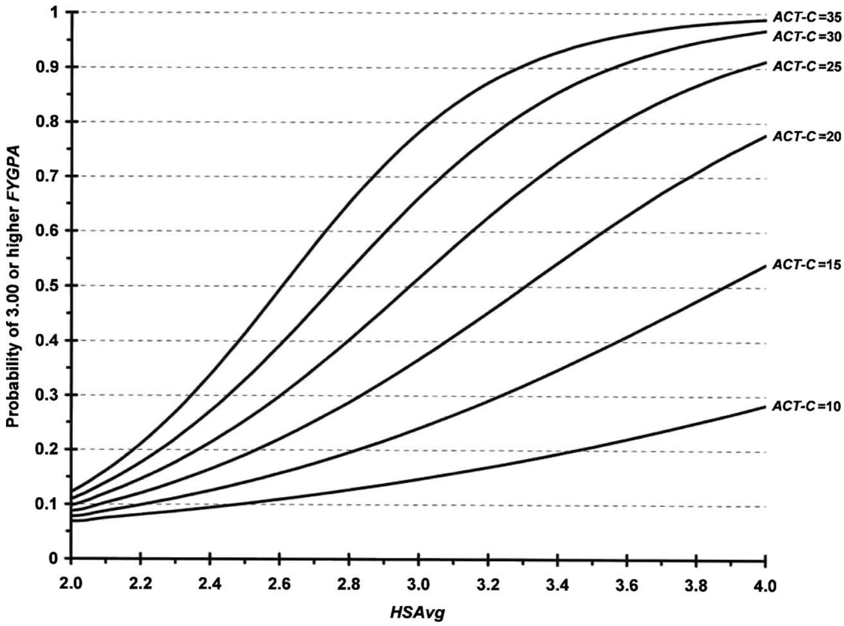

FIGURE 3 Probability of 3.00 or higher *FYGPA*, given *HSAvg* and *ACT-C* score.

In the joint models (labeled C in Table 2), the fixed effects for both the main effects and the interaction term are positive and statistically significant ($p < .001$). One interpretation of the interaction term is that *HSAvg* is more predictive for students with higher *ACT-C* scores than for students with lower *ACT-C* scores.[4] Figure 3 shows the probability of success (3.0 or higher), given different values of *HSAvg* and *ACT-C*. As *ACT-C* increases, the slope of the *HSAvg* probability-of-success curve increases markedly. Similar results occur for the other success levels.

The only statistically significant institution-level fixed effects in Table 2 are mean *HSAvg* and mean *ACT-C*. The coefficients associated with mean *HSAvg* as a predictor of the *HSAvg* slope are positive, which indicates that the *HSAvg* probability-of-success curves tend to be steeper at institutions where applicants have higher mean *HSAvg* than at institutions where applicants have lower mean *HSAvg*. In contrast, the slope coefficients associated with mean *ACT-C* as a predictor of the *ACT-C* slope are negative. This result indicates that the *ACT-C* probability-of-success curves are steeper at institutions where applicants have lower mean *ACT-C*. The other institution variables that I considered (affiliation, undergraduate enrollment, self-rated selectivity, and percent minority) did not meet the threshold of statistical significance ($p < .01$) required to enter the model after mean *HSAvg* and mean *ACT-C* had already been included. Apparently,

[4]Alternatively, one could say that *ACT-C* is more predictive for students with high *HSAvg* than for students with low *HSAvg*.

the effects of affiliation, percent minority, and undergraduate enrollment on the probability-of-success curves are not distinguishable from those associated with mean *HSAvg* and mean *ACT-C*.

Incremental Success Rate With Respect to Base Rate

Figure 4 shows median incremental success rates (Equation 8) associated with the 3.0 or higher success level and the three sets of selection variables. The solid curves in Figure 4 illustrate results calculated from the hierarchical models. *HSAvg* has higher incremental success rates than *ACT-C* at low to moderate cutoff proportions, but *ACT-C* does better than *HSAvg* at high cutoff proportions. At higher cutoff proportions, selection based on *ACT-C* and *HSAvg* jointly increases the incremental success rate over that for selection based on *HSAvg* only; this occurs around the cutoff proportion .40 (*HSAvg* = 3.2 or *ACT-C* = 19). Although not shown in Figure 4, similar results occurred for the other success levels.

The dashed curves in Figure 4 show success rates calculated from Equation 3 using correlation coefficients equal to the median correlations over all institutions. It is clear that the success rates based on an assumption of bivariate normality differ substantially from those modeled from the data. The bivariate normal success rate for *HSAvg* is smaller than the modeled success rate when *HSAvg* is less than 3.9, and often substantially so. On the other hand, the bivariate normal success rate for *HSAvg* is much larger than the modeled success rate when *HSAvg* is greater than 3.9. In contrast, the bivariate normal assumption results in underestimated success rates for all values of *ACT-C* and of the joint *HSAvg & ACT-C* selection variable.

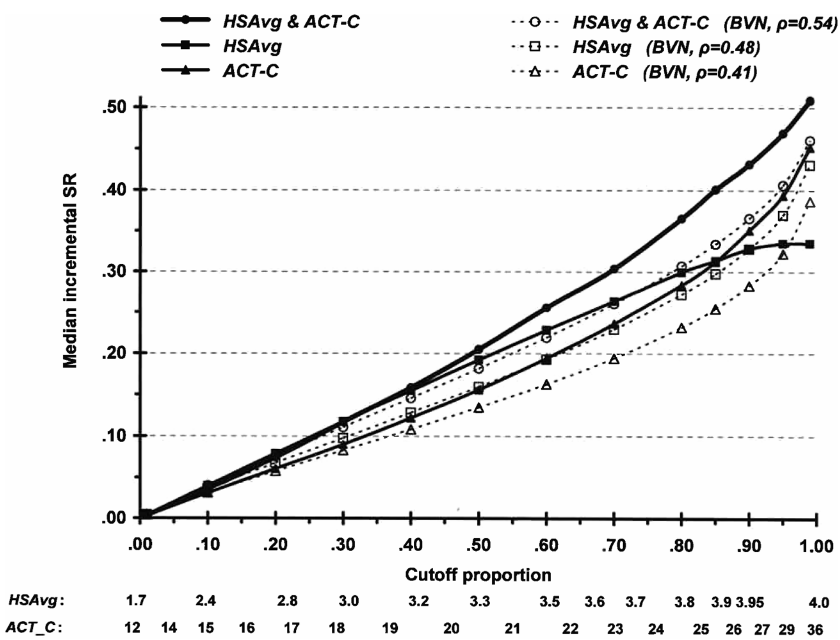

FIGURE 4 Median incremental success rate with respect to base success rate, by prediction model and cutoff proportion (3.0 or higher *FYGPA*).

Incremental Success Rate of *ACT-C* With Respect to *HSAvg*

Table 3 shows that the median incremental success rate of *ACT-C* with respect to *HSAvg* (Equation 9) depends on both success level and on cutoff proportion. For the 2.0 success level, *ACT-C* increases success rate only modestly above that attainable with *HSAvg*. As success level increases, the incremental success rate associated with *ACT-C* increases sharply at higher cutoff proportions.

The second column of Table 3 shows the approximate values of *HSAvg* associated with the cutoff proportions in the first column. This column suggests that for the 2.0 and 3.0 success levels, *ACT-C* typically has incremental usefulness with respect to *HSAvg* when *HSAvg* is 3.5 or greater. For the 3.5 and 3.7 success levels, *ACT-C* has incremental usefulness when *HSAvg* is at least 3.7 and 3.8, respectively.

Incremental Accuracy Rate with Respect to Null Decisions

Figure 5 shows the percentage of institutions for which there is incremental accuracy in selection at particular cutoff proportions (Equation 11). For the 2.0 success level, *HSAvg* has incremental accuracy at a majority of institutions only for very low cutoff proportions. *ACT-C* by itself does not have incremental accuracy at most institutions for any cutoff proportion. Both selection variables have incremental accuracy most frequently with respect to the 3.0 success level, which corresponds to typical achievement. For the 3.5 success level, both selection variables have incremental accuracy at a majority of institutions for high cutoff proportions. For the 3.7 success level, *ACT-C* and the joint model have incremental accuracy at a majority of institutions for very high cutoff proportions. *HSAvg*, in contrast, does not have incremental accuracy at most institutions for any cutoff proportion.

TABLE 3
Median Incremental Success Rate of *ACT-C* With Respect to *HSAvg*, by First-Year GPA Success Level and Cutoff Proportion (N = 192)

		Success Level			
Cutoff Proportion	*Approx. Value of HSAvg*	*2.0*	*3.0*	*3.5*	*3.7*
.01	1.7	.00	.00	.00	.00
.10	2.4	.00	.00	.00	.00
.20	2.7	.00	.00	.00	.00
.30	3.0	.00	.00	.00	.00
.40	3.2	.00	.00	.00	.00
.50	3.3	.00	.00	.00	.00
.60	3.5	.01	.01	.00	.00
.70	3.7	.01	.03	.01	.00
.80	3.8	.02	.05	.04	.03
.85	3.9	.02	.07	.07	.05
.90	3.95	.02	.09	.11	.09
.95	4.0	.03	.12	.17	.18
.99	4.0	.03	.16	.27	.30

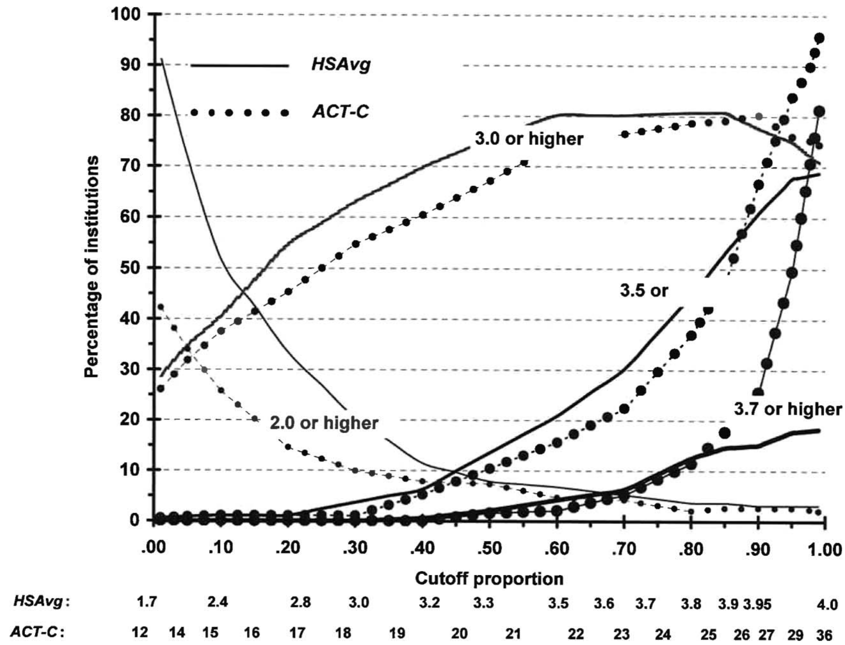

FIGURE 5 Percentage of institutions for which selection variables have incremental accuracy with respect to null decisions, by cutoff proportion.

Incremental Accuracy Rate of *ACT-C* With Respect to *HSAvg*

As shown in Table 4, *ACT-C* has little incremental accuracy with respect to *HSAvg* for the 2.0 success level, but it increases with success level. The maximum increase in accuracy rate associated with adding *ACT-C* to a selection rule based on *HSAvg* is about .04. The second column of Table 4 suggests that for the 3.0, 3.5, and 3.7 success levels, *ACT-C* typically has incremental accuracy with respect to *HSAvg* when *HSAvg* is 3.3 or greater.

SUMMARY AND DISCUSSION

The usefulness of selection variables depends on the goals intended by decision makers. In college admission, one plausible goal is to maximize the success rate (*SR*) of enrolled students with respect to some near-term variable (e.g., first-year GPA) as a prelude to longer-term academic success. Another plausible goal is to maximize the accuracy rate (*AR*) of an institution's decisions to admit or reject applicants. When the selection variable and outcome variable have a bivariate normal distribution with correlation coefficient ρ, the *SR* and *AR* indicators of usefulness are functions of the success level, the cutoff proportion (proportion of applicants not admitted), and ρ. It is perhaps for this reason that people often interpret ρ as an indicator of usefulness. The bivariate normal assumption is not realistic in college admission, however, because *HSAvg* has

TABLE 4
Median Incremental Accuracy Rate of *ACT-C* With Respect to *HSAvg*, by First-Year GPA Success Level and Cutoff Proportion

Cutoff Proportion	*Approx. Value of HSAvg*	*Success Level* 2.0	3.0	3.5	3.7
.01	1.7	.00	.00	.00	.00
.10	2.4	.00	.00	.00	.00
.20	2.7	.00	.00	.00	.00
.30	3.0	.00	.00	.00	.00
.40	3.2	.00	.00	.00	.00
.50	3.3	.00	.01	.02	.01
.60	3.5	.01	.02	.03	.02
.70	3.7	.00	.03	.03	.03
.80	3.8	.00	.03	.04	.04
.85	3.9	.01	.03	.04	.04
.90	3.95	.01	.03	.03	.04
.95	4.0	.00	.02	.02	.03
.99	4.0	.00	.01	.01	.01

a pronounced skew. This study suggests that *SR* and *AR* can depart substantially from values calculated from ρ.

Both *HSAvg* and *ACT-C* predict academic success in the first year of college. As shown in Figure 2, however, their probability-of-success curves vary with different levels of success and with different cutoff proportions. Both variables have steeper slopes for the higher levels of success (*S30, S35,* and *S37*) than for the minimal level of success (*S20*). *HSAvg* is a stronger predictor than *ACT-C* for *S20*, but *ACT-C* is much stronger than *HSAvg* for predicting *S35* and *S37*. *HSAvg* does better at lower cutoff proportions, and *ACT-C* does better at higher cutoff proportions.

The statistical relationship of *HSAvg* and *ACT-C* with any level of success also depends on the joint values of both predictors (Figure 3). *HSAvg* is a much stronger predictor among students with high *ACT-C* scores than among students with low *ACT-C* scores. Correspondingly, *ACT-C* is a much stronger predictor among students with high values of *HSAvg* than among students with low values of *HSAvg*.

There is moderate variation among institutions in all the intercept and slope coefficients defining the conditional probability of success functions. Typically, the standard deviations of the coefficients among institutions are about two-tenths to three-tenths of the corresponding mean values. This variation indicates that institutions would benefit from doing their own local predictive validity studies, rather than relying on global studies.

Institutions can use either *HSAvg* or *ACT-C* to increase their success rates beyond the base success rate, no matter which success level they choose. *HSAvg* is more effective than *ACT-C* for increasing success rates at low to moderate cutoff proportions, but *ACT-C* is more effective at higher cutoff proportions. Using both selection variables, however, is more beneficial for improving success rates than using either variable by itself for cutoffs above *HSAvg* = 3.5 or *ACT-C* = 22.

Increasing accuracy rates, beyond that associated with the null decisions of either admitting all applicants or denying admission to all applicants, is more difficult than increasing success rates.

The reason is that accurate classification requires both the success of admitted applicants and the failure of non-admitted applicants. The effectiveness of both *HSAvg* and *ACT-C* for increasing accuracy rates depends strongly on the success level.

These results are based on the assumption that institutions use strict cutoff proportions (c) in making admission decisions. This assumption is rarely if ever true, but it is necessary for this research, given the absence of quantifiable rules and data on institutions' actual admission policies. Institutions use other variables, in addition to high school course work, grades, and test scores, in making admission decisions. One should therefore interpret the results of this study with this in mind.

With this limitation in mind, we can conclude that the conventional wisdom based on correlations is correct in some, but not all respects. *HSAvg* by itself is better than *ACT-C* by itself for some, but not for all, degrees of selectivity and definitions of success. In some situations (e.g., where an institution is interested in high levels of success), *ACT-C* is more useful. This study affirms another aspect of the conventional wisdom, however: In most scenarios, using both high school grades and test scores jointly is better than using either by itself. In using both variables, moreover, it is important to take into account the *HSAvg* by *ACT-C* interaction effect, as well as the main effects.

Finally, it is worthwhile to keep in mind that although increasing *SR* and *AR* in making admission decisions is a plausible goal at many institutions, it is not their only goal. Other goals, such as objectivity and uniformity in making admission decisions, and providing data for counseling, placement, and institutional self-study are also important, and both high school grades and test scores contribute in different ways to these goals.

REFERENCES

ACT. (1999). *Prediction research summary tables*. (Available from ACT, Inc., P. O. Box 168, Iowa City, IA 52243).

ACT. (2008a). *The ACT user handbook*. Iowa City, IA: Author. Retrieved from http://www.act.org/aap/pdf/ACT_User_Handbook.pdf

ACT. (2008b). *Updated validity statistics for ACT Prediction Service*. (Available from ACT, Inc., P. O. Box 168, Iowa City, IA 52243).

Allen, J., Robbins, S., Casillas, A., & Oh, I.-S. (2008). Third-year college retention and transfer: Effects of academic performance, motivation, and social connectedness. *Research in Higher Education*, *49*(7), 647–664.

Bowen, W., Chingos, M., & McPherson, M. (2009). *Crossing the finish line: Completing colleges at America's public universities*. Princeton, NJ: Princeton University Press.

Breland, H., Maxey, J., Gernand, R., Cumming, T., & Trapani, C. (2002). *Trends in college admissions: A report of a survey of undergraduate admissions policies, practices, and procedures*. Association for the Institutional Research. Retrieved from http://www.airweb.org/images/trendsreport.pdf/

Camara, W. (2005). Broadening criteria of college success and the impact of cognitive predictors. In W. Camara & E. Kimmel (Eds.), *Choosing students: Higher education admission tools for the 21st century* (pp. 53–80). Mahwah, NJ: Lawrence Erlbaum Associates.

Casillas, A., Robbins, S. B., Allen J., Kuo, Y. K., Hanson, M. A., & Schmeiser, C. (2010). *Predicting early academic failure in high school from prior academic achievement, psychosocial characteristics, and behavior*. (Available from ACT, Inc., P. O. Box 168, Iowa City, IA).

Cronbach, L. J., & Gleser, G. C. (1965). *Psychological tests and personnel decisions*. Urbana, IL: University of Illinois Press.

Geisinger, K. (June, 2009). A college admission question: What would we do if the ACT and SAT did not exist? *NCME newsletter, 17* (2). Retrieved from http://www.ncme.org/pubs/pdf/vol_17_num_2.pdf

Kane, M. T. (2006). Validity. In R. L. Brennan (Ed.), *Educational measurement* (4th ed.; pp. 17–64). Westport, CT: Praeger Publishers.

Kobrin, J. L., & Patterson, B. F. (2010, April). *Exploring the variability in the validity of SAT scores for predicting first-year college grades at different colleges and universities*. Paper presented at the meeting of the American Educational Research Association, Denver, CO.

Kobrin, J. L., Patterson, B. F., Shaw, E. J., Mattern, K. D., & Barbuti, S. M. (2008). *Validity of the SAT for predicting first-year college grade point average*. New York, NY: College Board. Retrieved from http://professionals.collegeboard.com/data-reports-research/cb/validity-of-sat-predicting-fycgpa

Lenore-Jenkins, S., & Goff, J. (2008, September). *Admission vs. enrollment management: Separate but equal?* Paper presented at the meeting of the National Association for College Admission Counseling, Arlington, VA. Retrieved from http://www.nacacnet.org/EventsTraining/NC10/2008NationalConference/Documents/2008%20Educational%20Session%20Materials/B200.pdf

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.

National Association for College Admission Counseling. (2006). *The changing landscape of higher education: College admission in context*. Retrieved from http://www.nacacnet.org/PublicationsResources/Marketplace/research/Pages/CollegeAdmissioninContext.aspx

Noble, J. P., & Sawyer, R. L. (2004). Is high school GPA better than admission test scores for predicting academic success in college? *College and University Journal*, *79*(4), 17–22.

Perfetto, G. (1999). *Toward a taxonomy of the admission decision-making process: A public document based on the first and second College Board Conferences on Admissions Models*. New York, NY: The College Board. Retrieved from http://professionals.collegeboard.com/profdownload/pdf/gataxonomy030514_23496.pdf

Petersen, N. S., & Novick, M. R. (1976). An evaluation of some models for culture-fair selection. *Journal of Educational Measurement*, *13*, 3–29.

Raudenbush, S., Bryk, A., Cheong, Y. F., & Congdon, R. (2004). *HLM6: Hierarchical linear and nonlinear modeling*. Lincolnwood, IL: Scientific Software International.

SAS Institute. (2008). *The NLMIXED Procedure* (SAS/STAT User's Guide). Cary, NC: SAS Institute. Retrieved from http://support.sas.com/onlinedoc/913/docmainpage.jsp

Saupe, J. L., & Curs, B. R. (2008). Driving enrollment management scores from ACT data. *IR Applications*, *16*(15). Retrieved from http://www.airweb.org/Education/Events/Publications/ProfessionalFiles/iTapps16.pdf

Sawyer, R. (2007). Indicators of usefulness of test scores. *Applied Measurement in Education*, *20*(3), 255–271.

Sawyer, R. L. (1996). Decision theory models for validating course placement tests. *Journal of Educational Measurement*, *33*(3), 271–290.

Willingham, W. (2005). Prospects for improving grades for use in admissions. In W. Camara & E. Kimmel (Eds.), *Choosing students: Higher education admission tools for the 21st century* (pp. 127–140). Mahwah, NJ: Lawrence Erlbaum Associates.

Woodruff, D. J., & Ziomek, R. L. (2004). *High school grade inflation from 1991 to 2003*. Iowa City, IA: ACT. Retrieved from http://www.act.org/research/researchers/reports/pdf/ACT_RR2004-4.pdf

APPENDIX

Relationship of Conditional Probability of Success and Its Marginal Distribution With Success Rate and Accuracy Rate

Proposition 1:

Let $p(x) = P\,[\text{Success}|\text{X} = x]$ be strictly increasing in x
$f(x) =$ be the probability density function for X.

Assume that f is non-zero everywhere.

Let $SR(k) = \int_k^\infty p(x) f(x)\, dx \Big/ \int_k^\infty f(x)\, dx$ be the success rate for a cutoff score k, and let $BSR = SR(-\infty)$. Then:

1. *SR is strictly increasing in x.*
2. *SR(x) > BSR.*
3. *SR(x) > p(x).*

Proof:

$$\text{Let } SR(x) = \frac{\int_x^\infty p(t)f(t)dt}{\int_x^\infty f(t)dt} = \frac{A}{B}.$$

$$\begin{aligned}\text{Then } SR'(x) &= \frac{A'B - AB'}{B^2}\\ &= \frac{-f(x)p(x)B + Af(x)}{B^2}\\ &= \frac{f(x)\left[A - Bp(x)\right]}{B^2}.\end{aligned}$$

$$\begin{aligned}\text{Since } p \text{ is increasing, } A &= \int_x^\infty p(t)f(t)dt > p(x)\int_x^\infty f(t)dt\\ &= p(x)B,\end{aligned}$$

and therefore, $SR'(x) > 0$. Thus, SR is strictly increasing in x, and $SR(x) > SR(-\infty) = BSR$. Because $\frac{f(x)\left[A - Bp(x)\right]}{B^2} > 0$ for all x, $p(x)\int_x^\infty f(t)dt < \int_x^\infty p(t)f(t)dt$ for all x, and so $SR(x) > p(x)$ for all x.

Proposition 2:

Given the previous assumptions, let $AR\,(k) = \int_{-\infty}^k \left[1 - p\,(x)\right] f\,(x)\, dx + \int_k^\infty p\,(x) f\,(x)\, dx$ be the accuracy rate associated with cutoff score k. Then, AR has a local maximum at k_0 if, and only if, $p\,(k_0) = 1/2$.

Proof:

$$AR'(k) = \left[1 - p(k)\right] f(k) - p(k) f(k)$$
$$= f(k) - 2p(k) f(k) .$$

$AR'(k_0) = 0$ if, and only if, $-2p(k_0) f(k_0) = f(k_0)$.

Since $f(k_0) > 0$, $AR'(k_0) = 0$ if, and only if, $p(k_0) = 1/2$.

Note that $AR''(k) = f'(k) - 2\left[p'(k)f(k) + p(k) f'(k)\right]$.

Therefore $AR''(k_0) = f'(k_0) - 2p'(k_0)f(k_0) - f'(k_0)$

$$= -2p'(k_0)f(k_0), \text{ since } p(k_0) = 1/2.$$

Because $p'(k_0) > 0$ and $f(k_0) > 0$, $AR''(k_0) < 0$.

Therefore, AR has a local maximum at k_0 if, and only if, $p(k_0) = 1/2$.

Proposition 3:

Let $BSR = \int_{-\infty}^{\infty} p(x) f(x)\, dx$. Then max AR exceeds BSR and 1-BSR if, and only if, p crosses 1/2.

Proof:

Suppose $AR(k) > BSR$ for some k. Proof by contradiction:

Suppose $p(x) > 1/2$ for all x. Then $\int_{-\infty}^{k} \left[1 - p(x)\right] f(x)\, dx < \int_{-\infty}^{k} p(x) f(x)\, dx$.

And so, $AR(k) < \int_{-\infty}^{k} p(x) f(x)\, dx + \int_{k}^{\infty} p(x) f(x)\, dx$

$$= BSR \text{ [contradiction]}.$$

If $p(x) < 1/2$ for all x, then $\int_{k}^{\infty} p(x) f(x)\, dx < \int_{k}^{\infty} \left[1 - p(x)\right] f(x)\, dx$.

And so, $AR(k) < \int_{-\infty}^{k} \left[1 - p(x)\right] f(x)\, dx + \int_{k}^{\infty} \left[1 - p(x)\right] f(x)\, dx$

$$= \int_{-\infty}^{\infty} \left[1 - p(x)\right] f(x)\, dx$$

$$= 1 - BSR[\text{contradiction}].$$

Therefore $p(k_0) = 1/2$ for some k_0.

Since p is monotonic increasing, p must cross $1/2$ at k_0.

Now, suppose p crosses $1/2$. Let $p(k_0) = 1/2$ and $p(x) < 1/2$, for $x < k_0$, and $p(x) > 1/2$, for $x > k_0$.

$$\text{Then } AR(k_0) = \int_{-\infty}^{k_0} \left[1 - p(x)\right] f(x) dx + \int_{k_0}^{\infty} p(x) f(x)\, dx$$
$$> \int_{-\infty}^{k_0} p(x) f(x)\, dx + \int_{k_0}^{\infty} p(x) f(x)\, dx$$
$$= BSR\,.$$

$$\text{Also, } AR(k_0) = \int_{-\infty}^{k_0} \left[1 - p(x)\right] f(x)\, dx + \int_{k_0}^{\infty} p(x) f(x)\, dx$$
$$> \int_{-\infty}^{k_0} \left[1 - p(x)\right] f(x)\, dx + \int_{k_0}^{\infty} \left[1 - p(x)\right] f(x)\, dx$$
$$= 1 - BSR\,.$$