



Research Report

ETS RR-13-09

Disentangling the Role of High School Grades, SAT[®] Scores, and SES in Predicting College Achievement

Rebecca Zwick

May 2013

ETS Research Report Series

EIGNOR EXECUTIVE EDITOR

James Carlson
Principal Psychometrician

ASSOCIATE EDITORS

Beata Beigman Klebanov
Research Scientist

Heather Buzick
Research Scientist

Brent Bridgeman
Distinguished Presidential Appointee

Keelan Evanini
Research Scientist

Marna Golub-Smith
Principal Psychometrician

Shelby Haberman
Distinguished Presidential Appointee

Gary Ockey
Research Scientist

Donald Powers
Managing Principal Research Scientist

Frank Rijmen
Principal Research Scientist

John Sabatini
Managing Principal Research Scientist

Matthias von Davier
Director, Research

Rebecca Zwick
Distinguished Presidential Appointee

PRODUCTION EDITORS

Kim Fryer
Manager, Editing Services

Ruth Greenwood
Editor

Since its 1947 founding, ETS has conducted and disseminated scientific research to support its products and services, and to advance the measurement and education fields. In keeping with these goals, ETS is committed to making its research freely available to the professional community and to the general public. Published accounts of ETS research, including papers in the ETS Research Report series, undergo a formal peer-review process by ETS staff to ensure that they meet established scientific and professional standards. All such ETS-conducted peer reviews are in addition to any reviews that outside organizations may provide as part of their own publication processes. Peer review notwithstanding, the positions expressed in the ETS Research Report series and other published accounts of ETS research are those of the authors and not necessarily those of the Officers and Trustees of Educational Testing Service.

The Daniel Eignor Editorship is named in honor of Dr. Daniel R. Eignor, who from 2001 until 2011 served the Research and Development division as Editor for the ETS Research Report series. The Eignor Editorship has been created to recognize the pivotal leadership role that Dr. Eignor played in the research publication process at ETS.

**Disentangling the Role of High School Grades, SAT[®] Scores, and SES
in Predicting College Achievement**

Rebecca Zwick
ETS, Princeton, New Jersey

May 2013

Find other ETS-published reports by searching the ETS
ReSEARCHER database at <http://search.ets.org/researcher/>

To obtain a copy of an ETS research report, please visit
<http://www.ets.org/research/contact.html>

Action Editor: Marna Golub-Smith

Reviewers: Brent Bridgeman and Sandip Sinharay

Copyright © 2013 by Educational Testing Service. All rights reserved.

ETS, the ETS logo, and LISTENING. LEARNING. LEADING. are
registered trademarks of Educational Testing Service (ETS).

SAT is a registered trademark of the College Board.

PSAT/NMSQT is a registered trademark of the College Board and the National Merit Scholarship Corporation.



Abstract

Focusing on high school grade-point average (HSGPA) in college admissions may foster ethnic diversity and communicate the importance of high school performance. It has further been claimed that HSGPA is the best single predictor of college grades and that it is more equitable than test scores because of a smaller association with socioeconomic status (SES). Recent findings, however, suggest that HSPGA's seemingly smaller correlation with SES is a methodological artifact. In addition, it tends to produce systematic errors in the prediction of college grades. Although supplementing HSGPA with a high-school resource index can mitigate these errors, determining whether to include such an index in admissions decisions must take into account the institutional mission and the potential diversity impact.

Key words: college admissions, socioeconomic status, *SAT*[®], high school grades

Acknowledgments

I am grateful for the support of the College Board, which provided data for the Zwick and Green (2007) and Zwick and Himelfarb (2011) studies and provided funding for the Zwick and Himelfarb study. Both studies were conducted at the University of California, Santa Barbara. This paper benefitted from comments by Brent Bridgeman, Marna Golub-Smith, and Sandip Sinharay.

Attempts to statistically forecast college grades using high school performance and admissions test scores have a long history, dating back at least as far as 1939, when Theodore Sarbin compared predictions of first-quarter college grades obtained via regression analysis to predictions based on counselor judgment (Sarbin, 1943; the regression equation won, albeit by a small amount.) Since that time, thousands of statistical studies have been conducted in which the goal was to predict college grade-point average using test scores (nearly always the *SAT*[®] or ACT) and *high school grade-point average* (HSGPA).

These studies fall into two categories. First, many undergraduate institutions conduct research to predict students' college performance using various combinations of possible admissions criteria (e.g., Geiser, 2004). For reasons of convenience, the most commonly used measure of college performance is *first-year grade-point average* (FGPA). Second, testing companies conduct similar studies to investigate the validity of admissions tests as predictors of college achievement (e.g., Kobrin, Patterson, Shaw, Mattern, & Barbuti, 2008). These two kinds of studies typically assess the predictive value of both HSGPA and admissions test scores because these are the academic indicators most often used in the college admissions process.

The resulting body of research literature has produced one remarkably consistent belief: High school grade-point average is the best predictor of college grade-point average. For example, the introduction to the new book, *SAT Wars*, states, "Lest there be any confusion about this, one should keep in mind that high school grade-point average (HSGPA) has always been the best single academic variable predicting college grades ..." (Soares, 2012, p. 6). And in their essay on college admissions testing, Atkinson and Geiser (2009, p. 666) asserted that "it is useful to begin any discussion of standardized admissions tests with acknowledgment that a student's record in college preparatory courses in high school remains the best indicator of how the student is likely to perform in college."

In fact, the determination of whether high school grades are the optimal predictor of college performance may depend on the evaluation criteria. One way to compare the quality of predictors is in terms of how highly correlated they are with the target of prediction—in this case, college grades. (A similar comparison can be conducted in terms of the R^2 values from a regression analysis, as explained below.) From this perspective, high school GPA tends to make a relatively strong showing. For example, in a widely publicized study of the SAT at the University of California, Geiser (2004) found the squared correlation between HSGPA and

FGPA to be .15, compared to only .13 for SAT scores. Yet it can be misleading to evaluate a predictor simply in terms of correlations. A second aspect of prediction quality that is often overlooked involves systematic prediction errors that can exist for some student groups, even in the presence of a strong association between the predictor and the predicted quantity. With respect to this criterion, HSGPA does not perform well as a predictor. A third criterion that has emerged may be more political than technical: Some commentators argue that predictors of college performance should be unassociated with socioeconomic status (SES) to the degree possible. From this perspective, HSGPA has been deemed a superior predictor to test scores because it is said to be relatively untainted by SES and thus a more legitimate predictor of FGPA. For example, Geiser and Santelices (2007) asserted that “[c]ompared to high-school grade-point average (HSGPA), scores on standardized admissions tests such as the SAT I [an earlier name for the SAT] are much more closely correlated with students’ socioeconomic background characteristics” (p. 2) and use this claim to support their conclusion that “high-school grades provide a fairer, more equitable and ultimately more meaningful basis for admissions decision-making...” (p. 27). Their discussion is in keeping with the belief, expressed in some recent critiques of testing, that correlations between admissions test scores and subsequent grades are largely, if not entirely, “an artifact of the common influences of SES on both test scores and grades” (see Sackett, Kuncel, Arneson, Cooper, & Waters, 2009, p. 2 for a discussion).¹ Again, however, a more fine-grained analysis yields a different conclusion.

In the following section, the findings of several large studies that sought to predict college grades using test scores and HSGPA are described. The discussion here is restricted to the SAT, but is largely applicable to the ACT as well. The third section summarizes the results of a study that illustrates the substantial prediction errors that result for some ethnic groups when HSGPA alone is used as a predictor of college performance. These errors are mitigated by the inclusion of an index intended to serve as a proxy for school resources and also by the inclusion of SAT scores. The fourth section describes a study that calls into question the claims about the degree to which SAT scores and HSGPA are associated with SES. The results suggest that test scores and high school grades have similar correlations with SES at the student level. These findings are pertinent to the role of HSGPA in predicting FGPA. However, as addressed in the *Discussion* section, predicted college GPA need not be the basis for college admissions

decisions. In fact, predicted performance need not play any role at all. Admissions criteria must be established so as to support the goals of the institution.

Role of High School GPA in Equations for Predicting College GPA

Studies that assess the strength of association between academic indicators (such as standardized admissions test scores and HSGPA) and college success at undergraduate institutions are typically conducted by performing a regression analysis in which SAT scores and HSGPA are used to predict first-year FGPA for students who already have FGPA's on record. (Whether FGPA is the optimal college performance measure is, of course, debatable. Some studies have used college course grades, later college GPAs, or degree attainment as success criteria. See, e.g., Burton & Ramist, 2001; Geiser & Santelices, 2007.) The correlation between the predicted FGPA and the actual FGPA earned by the students can then be estimated. This correlation (R), or more commonly, its square (R^2), is used as a measure of the strength of prediction. The strength of prediction can also be evaluated for equations that include HSGPA only or admissions test scores only so that these competing models can be compared. The squared correlation for each model can be interpreted as the proportion of variance in FGPA that is attributable to, or explained by the predictors. These analyses not only produce results that are useful in evaluating the various prediction models, but also yield equations that can be used to predict the performance of applicants.

Table 1 gives the relevant squared correlations for three large studies that have appeared in recent years. There are several noteworthy aspects to the results. First, the joint contribution of HSGPA and SAT is less than the sum of their individual contributions because the predictors are not independent of each other. Second, both the results for SAT only and the results for HSGPA and SAT combined are nearly identical across the three studies, despite the differences among the studies in terms of the samples of students, the time of data collection, and other methodological factors. Twelve to 13% of the variance in FGPA was attributable to SAT scores; 21 to 22% of the variance was explained by HSGPA and SAT scores combined. The third notable finding in the table is that the squared correlation for HSGPA only is always larger than the squared correlation for SAT alone (although the disparity is large only in the Zwick and Sklar (2005) study, which was based on data that are now roughly three decades old). This recurrent finding is the basis for the claim that high school grades are the best predictor of college grades.

Table 1***Squared Correlations (R^2) Between Predictors and First-Year College GPA in Three Studies***

| | Kobrin et al. (2008, p. 5) | Geiser (2004, p. 130) | Zwick & Sklar (2005, p. 451) |
|-----------------------|--|---|--|
| HSGPA only | .13 | .15 | .21 |
| SAT only | .12 | .13 | .12 |
| HSGPA plus SAT | .21 | .21 | .22 |
| Nature of sample | Students in the fall 2006 entering cohort at 110 institutions around the United States | Students who entered seven campuses of the University of California in 1996–1999. (UC Riverside entrants for 1997 and 1998 were excluded because of data problems. UC Santa Cruz was excluded because it did not provide conventional grades during this period.) | Students in the 1992 follow-up of the High School and Beyond (HSB) 1980 sophomore cohort, a national probability sample |
| Number of students | 151,316 | 77,893 | 4,617 |
| Details on HSGPA | HSGPA was self-reported on the SAT Questionnaire in terms of 12 ordered categories. | HS grades were from transcripts. HSGPA was the honors-weighted GPA used by UC at the time of the study: Additional grade points were included for honors courses. | HSGPA was from transcripts. |
| Details on SAT scores | SAT Math, Critical Reading, and Writing scores (components of the SAT since 2005) were included separately in the regressions. | SAT score was the sum of SAT Math and Verbal scores. | In the HSB database, SAT score was the sum of Math and Verbal scores. For students who took only the ACT or <i>PSAT/NMSQT</i> [®] , an SAT-equivalent score is substituted. |

Note. No corrections have been applied to the R^2 values. All studies excluded students with missing data. GPA = grade-point average; HSGPA = high school grade-point average.

It is worth noting that there have been exceptions to this pattern of superior predictive value for HSGPA. For example, in the most recent of the four University of California (UC)

entry cohorts studied by Geiser (2004, p. 130), SAT scores were found to be more predictive of FGPA than was HSGPA. In a later study of two UC entering classes, Agronow and Studley (2007) found that in 2006, but not 2004, the SAT showed a very small advantage over HSGPA in terms of predictive strength. Kobrin et al. (2008) found the SAT to be slightly more predictive than HSGPA in certain categories of institutions, including those with admission rates under 50%, and those with fewer than 2,000 undergraduates. But more significant than these exceptions is the fact that, as detailed in the next section, predictive value does not depend solely on strength of association.

Systematic Errors in Predicting FGPA

A different perspective on the relative utility of potential predictors of FGPA is obtained by examining the pattern of prediction errors for key student groups. Regression analysis yields a model-predicted FGPA for each student, along with the departure of this predicted value from the FGPA actually earned by the student, called the *prediction error* or *residual*. Although it is a property of ordinary least squares regression (the standard type of regression analysis) that these residuals must average to zero across all individuals in the analysis, this need not be true within key student groups. Even a strong prediction model can produce predicted FGPA's that are systematically too high or too low for certain groups.

Research conducted over the last 45 years has shown that, for African American and Latino students, FGPA tends to be overpredicted (e.g., Cleary, 1968; Linn, 1983). That is, these students are, on average, predicted to earn higher FGPA's than they actually do. This recurrent result has become more widely known in recent years, following the publication of *The Shape of the River* (Bowen & Bok, 1998) and *The Black-White Test Score Gap* (Jencks & Phillips, 1998). Young (2004, pp. 293–294) found that, on average, African American students were overpredicted by .11 on a 0–4 scale, based on 11 studies in which SAT scores and HSGPA were used to predict FGPA. Latino students were overpredicted by an average of .08, based on eight studies. Identical prediction errors were obtained by Mattern, Patterson, Shaw, Kobrin, and Barbuti (2008, p. 11), who analyzed data from more than 150,000 students who took the SAT in 2006. A number of conjectures have been offered about the reasons for the persistent overprediction for African American and Latino students, including hypotheses involving inhospitable campus environments, negative student attitudes, and measurement error in HSGPA and admissions test scores (see Zwick, 2002, pp. 117–124; Zwick & Himelfarb, 2011).

An aspect of the ethnic-group overprediction phenomenon that is less well known is that this overprediction tends to be more severe when HSGPA alone is used as a predictor than when HSGPA and SAT scores are used in conjunction (e.g., see Mattern et al., 2008; Zwick & Schlemer, 2004).² A plausible hypothesis, investigated by Zwick and Himelfarb (2011), is that the overprediction of Latino and African American college grades occurs, at least in part, because these students are more likely than their White counterparts to attend high schools with fewer well-trained teachers and fewer instructional resources. Why might prediction errors be more severe when HSGPA alone is used as a predictor? High schools differ widely in their grading standards. Because high school teachers are concerned with ranking students within and not across schools, they can be expected to assign the entire range of grades. They are unlikely to refrain from assigning high grades merely because other high schools are viewed as being superior. This leads to a situation in which high schools tend to have similar distributions of grades, regardless of their relative achievement levels. Therefore, Zwick and Himelfarb (2011) conjectured that

using HSGPA alone as a predictor is expected to yield a misleadingly high predicted FGPA for students attending low-quality high schools. This overprediction is likely to be mitigated (but not eliminated) by the inclusion of SAT scores in the prediction equation because SAT scores have essentially the same meaning across schools. Under this hypothesis, the typical pattern of ethnic-group prediction errors occurs in part because ethnicity is, in effect, serving as a rough proxy for high school quality. (p. 103)

Table 2 shows the patterns of residuals for several alternative models for predicting FGPA. The table is based on analyses of data from 70,712 students conducted by Zwick and Himelfarb (2011). The data, provided by the College Board, were collected as part of an SAT validity study (Kobrin et al., 2008) that involved students who enrolled in 110 undergraduate institutions in 2006. Data from 34 public institutions proved to be suitable for the analysis. Students in the sample had attended a total of 5,702 high schools. Regressions were conducted within each college and the resulting residuals were then pooled for further analysis. The residuals are on the FGPA scale, ranging from 0 to 4.

First consider the results for Model 1, the typical model used to predict FGPA, which includes SAT Math, Critical Reading, and Writing scores, along with HSGPA. For White and Asian American students and those of Other ethnicity (which includes those who declined to

give their ethnicity), prediction error is very small—.01 or .02 on the FGPA scale.³ For groups that are underrepresented in higher education—African American, Latino, and Native American students—average prediction errors range from -.08 to -.12, meaning that predicted FGPA's are .08 to .12 higher than actual FGPA's, on average. The next column shows the average residuals for Model 2, in which HSGPA only is used to predict FGPA. For White, Asian American, and Other students, the average prediction error is, again, small—.03 for each of these groups. But for African American students, the average error is now -.25, meaning that African Americans tended to obtain FGPA's that were lower than predicted by a quarter of a grade point. (In one college in the study, the average prediction error for African American students reached three-quarters of a grade point in Model 2.) For Latino students, the average prediction error across all 34 colleges is a fifth of a grade point. The magnitude of these prediction errors when only HSGPA is used to predict FGPA, rarely mentioned in discussions of admissions criteria, clearly merits further attention.⁴

Zwick and Himelfarb (2011) sought to mitigate prediction errors by developing an index that would reflect the high school's instructional resources and including this index in prediction equations. Because of the unavailability of direct measures of school resources, they relied instead on measures of the average SES of the students attending the school, which they presumed to be related to the resources available to the school. The index they eventually created was defined as follows: high school SES index = district poverty level + average father education + average mother education. District poverty level was based on information on the poverty level of the district in which the high school was located. It was coded on a 1–4 scale, with 1 indicating a high poverty percentage and 4 indicating a low percentage. Average father education was the high-school-level average of student-reported father education (rated on a 1–9 scale, ranging from *grade school* to *graduate or professional degree*) and average mother education was defined in a corresponding fashion. (This represented a roughly equal weighting, because these three variables had roughly equal variances.) Each high school had a single value of the high school SES index; all students who attended the high school shared that same value.

Table 2***Average Residuals (Observed Minus Predicted) for Six Ethnic Groups Under Four Prediction Models for the Zwick and Himelfarb (2011) Data***

| Ethnic group | Sample size | % of group attending low-SES schools | Model 1 (HSGPA + SAT) ($R^2 = .21$) | Model 2 (HSGPA) ($R^2 = .15$) | Model 3 (HSGPA + SES index) ($R^2 = .17$) | Model 4 (HSGPA + SAT + SES index) ($R^2 = .23$) |
|------------------|-------------|--------------------------------------|---------------------------------------|---------------------------------|---|---|
| African American | 4,168 | 67 | -.12 | -.25 | -.17 | -.09 |
| Asian American | 6,939 | 47 | .02 | .03 | .01 | .02 |
| Latino | 5,087 | 72 | -.09 | -.20 | -.07 | -.03 |
| Native American | 387 | 59 | -.08 | -.10 | -.09 | -.08 |
| White | 49,671 | 47 | .02 | .03 | .02 | .01 |
| Other | 4,460 | 46 | .01 | .03 | .02 | .01 |

Note. Regressions were conducted within colleges. Residuals were then pooled for further analysis. Residuals are in terms of the FGPA scale (0–4). A minus sign indicates that on average, the predicted FGPA is higher than the actual FGPA. Approximate standard errors for average residuals in Model 2 (which has the smallest R^2) are .010 for African Americans, .008 for Asian Americans, .009 for Latinos, .033 for Native Americans, .003 for Whites, and .010 for Other. Standard errors are slightly smaller for the remaining models. The R^2 values are sample-size-weighted averages of the values from the 34 colleges. No corrections were applied to R^2 values. HSGPA = high school grade-point average; SES = socioeconomic status.

Zwick and Himelfarb (2011) sought to mitigate prediction errors by developing an index that would reflect the high school’s instructional resources and including this index in prediction equations. Because of the unavailability of direct measures of school resources, they relied instead on measures of the average SES of the students attending the school, which they presumed to be related to the resources available to the school. The index they eventually created was defined as follows: high school SES index = district poverty level + average father education + average mother education. District poverty level was based on information on the poverty level of the district in which the high school was located. It was coded on a 1–4 scale, with 1 indicating a high poverty percentage and 4 indicating a low percentage. Average father education was the high-school-level average of student-reported father education (rated on a 1–9 scale, ranging from *grade school* to *graduate or professional degree*) and average mother education was defined in a corresponding fashion. (This represented a roughly equal weighting, because

these three variables had roughly equal variances.) Each high school had a single value of the high school SES index; all students who attended the high school shared that same value.

Zwick and Himelfarb (2011) divided the nearly 71,000 students into two groups according to whether their high school was above or below the median value of the high school SES index. They found that, consistent with prediction, the students in the low-SES schools tended to receive predicted FGPA's that were higher (by an average of .10) than their actual grades, while the reverse was true for students in the high-SES schools: They performed better than predicted by an average of .10 (see Zwick & Himelfarb, 2011, Table 5). Also, as expected, African American and Latino students (who had the largest average overprediction errors) were much more likely to be in the low-SES high schools than were White, Asian American, and Other students (for whom prediction tended to be quite accurate). Column 3 of Table 2 shows that while 46 to 47% of White, Asian American, and Other students attended low-SES schools, 67% of African Americans and 72% of Latinos attended such schools. Fifty-nine percent of Native Americans (who tended to be overpredicted, but to a lesser degree than African American or Latino students) attended low-SES schools.

What happens when the high school SES index is added to the prediction model? Model 3 in Table 2 includes HSGPA and the index as predictors. Although it explains less variance than Model 1, the standard prediction model, it produces a similar pattern of residuals. An exception is that Model 1 tends to produce a more accurate prediction for African American students, with an average residual of -.12, compared to -.17 for Model 3. Model 4, which includes HSGPA, SAT, and the SES index, produces the most accurate pattern of average residuals, although it is only trivially better than Model 1 in terms of explained variance. In Model 4, all groups have an average prediction error of less than a 10th of a grade point.

The high school SES index Zwick and Himelfarb (2011) derived was limited in several respects. In particular, they did not have access to such instructionally relevant variables as teacher-student ratio or instructional expenditures. Also, average father education and average mother education were derived from test-takers' responses to the SAT Questionnaire and were probably not fully accurate. In addition, these high-school-level averages were based on only the students included in the dataset, rather than on all students attending the high school. Despite these limitations, the high school index did serve to reduce systematic errors in the prediction of

FGPA. It is likely that an index that included more instructionally relevant variables would produce a further reduction in prediction errors.

In short, these results supported the conjecture that, because African Americans, Latinos, and Native Americans were disproportionately represented in the low-SES high schools, their performance was, *on average*, overpredicted in Models 1 and 2 (Table 2). The overprediction apparently resulted in part from the fact that high school SES effects were masquerading as ethnicity effects. The addition of the high school SES index (Models 3 and 4) served to mitigate the systematic prediction errors in Model 2. This finding is consistent with the hypothesis that HSGPA tends to produce distorted predictions because of the differences across schools in the meaning of a particular grade. Evidently, even a rough measure of school resources can increase prediction accuracy. Whether or not it is desirable or fair to use such an index as part of admissions decisions is a more complex question that is addressed in the *Discussion* section.

Association Between FGPA Predictors and Socioeconomic Status

As noted earlier, another advantage that has been claimed for high school grades is that their predictive power is considered to be uncontaminated by effects of SES. But the association among SES, HSGPA, and SAT scores is a more complex phenomenon than it first appears, for reasons related to the same unusual properties of HSGPA that were discussed in the previous section: It characterizes students' accomplishments within a high school, but is not comparable across high schools. (A similar phenomenon applies to high school class rank, as illustrated below.) In a recent study that included 98,391 students and 7,330 high schools from across the nation, Zwick and Green (2007) found that within high schools, the association between high school grades and SES was not very different from the association between SAT scores and SES. A similar result was obtained by Willingham, Pollack, and Lewis (2002, p. 14) in their analysis of high school grades, SES, and test scores from the National Educational Longitudinal Study.

To understand the outcome of the Zwick and Green (2007) research, it is necessary to understand the differences among several types of correlations. Suppose we have the family income (as an SES indicator) and HSGPA for all N_S students in all N_H high schools in the United States. Now consider three types of correlations:

1. Correlation based on combined data: the kind of correlation that is typically reported, based on data consisting of a pair of values (income and HSGPA) for each of the N_S students, without taking into account which high school each student comes from.
2. Average within-high-school correlation: the value obtained by first calculating the ordinary correlation within each high school, and then averaging these correlations across the N_H high schools (taking the size of the high school into account).
3. Between-high-schools correlation: the correlation of high school average income with average HSGPA, based on data consisting of a pair of values (mean income and mean HSGPA) for each of the N_H high schools.

Each of these correlations can be illustrated using the artificial data in Table 3 (which expands an example from Zwick & Green, 2007). The table is intended to represent HSGPA, family income values (in thousands of dollars), and SAT scores (all three sections combined) from three high schools, each of which has six students. First, consider HSGPA and income only. The correlation obtained using the standard approach in (1)—combining the data for all 18 students and then computing the correlation—is only .096, close to zero. Yet within each of the high schools, there is a moderate positive association between HSGPA and income. In fact, the correlation within each school is exactly .378, so, of course, the average within-high-school correlation in (2) is also .378. The value of .378 represents the association between HSGPA and income at the student level. Why does the ordinary correlation in (1) give a misleading result? The surprisingly low value of .096 is obtained because the correlation in (1) reflects not only the average within-school association in (2) but also the between-schools association in (3). In this example, there is no relationship between a high school's average HSGPA and its average income: Although the three schools have different average incomes, they have the same average HSGPA. Therefore, the value of the correlation in (3) is zero. These correlations are displayed in Table 4. As shown algebraically in Zwick and Green (2007) and easily verified using Table 3, the average within-high-school correlation of .378 and the between-high-schools correlation of zero together yield a combined-data correlation of .096. Although this corresponds to the value that would ordinarily be reported in this situation, it is not a good reflection of the relationship between HSGPA and income at the student level.

Table 3***Artificial Data on HSGPA, Income, and SAT Score for Three High Schools***

| | High School A | | | High School B | | | High School C | | |
|---------|---------------|--------|------|---------------|--------|------|---------------|--------|------|
| | HSGPA | Income | SAT | HSGPA | Income | SAT | HSGPA | Income | SAT |
| Student | 1.000 | 25 | 1150 | 1.000 | 55 | 1350 | 1.000 | 85 | 1400 |
| data | 1.500 | 15 | 1450 | 1.500 | 45 | 1650 | 1.500 | 75 | 1700 |
| | 2.000 | 30 | 1500 | 2.000 | 60 | 1700 | 2.000 | 90 | 1750 |
| | 2.500 | 20 | 1800 | 2.500 | 50 | 2000 | 2.500 | 80 | 2050 |
| | 3.000 | 35 | 1850 | 3.000 | 65 | 2050 | 3.000 | 95 | 2100 |
| | 3.500 | 25 | 2150 | 3.500 | 55 | 2350 | 3.500 | 85 | 2400 |
| Mean | 2.250 | 25 | 1650 | 2.250 | 55 | | 2.250 | 85 | |

Note. Each high school has six students. Income = annual family income in thousands of dollars. HSGPA = high school grade-point average; SES = socioeconomic status. Table adapted from Table 2 of Zwick and Green (2007).

Table 4***Correlations Based on the Artificial Data of Table 3***

| Variables | Combined-data (ordinary) correlation | Average within-high-school correlation | Between-high-school correlation |
|-------------------|--------------------------------------|--|---------------------------------|
| HSGPA with income | .096 | .378 | 0 |
| SAT with income | .338 | .200 | .945 |

Note. HSGPA = high school grade-point average; SES = socioeconomic status.

Unlike HSGPA (and high school class rank), the high school averages of SAT scores do vary substantially between schools, and high school averages of SAT scores tend to be correlated with high school averages of family income. (This is the correlation in the between-high-school-correlation, above.) As shown in Table 4, it can be the case that the average within-high-school correlation of SAT and income is smaller than the combined-data correlation. For the data of Table 3, where the combined-data correlation is .338, the average within-high school correlation is .2, and the between-high-school correlation is .945 (Table 4). While the distinction among the three types of correlations presented here may seem unfamiliar, it is based on the principles of multilevel analysis. *Multilevel* (or *hierarchical*) *analysis*, which seeks to unconfound within-group and between-group phenomena, is widely recognized as the optimal analysis method for hierarchical data structures in which individuals are grouped into units (e.g., see Raudenbush &

Bryk, 2001). In practice, conducting analyses that take high school membership into account are not usually possible because information on high school membership is unavailable.

Now consider the relevant correlations for the Zwick and Green (2007) data, which are much less extreme than those of Tables 3 and 4. Table 5 highlights key results for White students, who comprised 64% of the sample. This is the group for which the research hypotheses were most clearly supported. Results for the entire sample and for each gender and ethnic group are given in Zwick and Green. The SES variables included in Table 5 are income and father's education; results for mother's education were very similar to those for father's education. (Income and parental education were reported by the student; SAT scores, from 2004, were from College Board records.) Table 5 shows that, for White students, the correlations of HSGPA with income and with father's education were slightly higher within schools than in the combined-data analysis, and this pattern was more pronounced for high school class rank. The reverse was true for SAT scores: Average within-high-school correlations were considerably smaller than combined-data correlations. As a result, HSGPA, class rank, and SAT had within-high school correlations with income and with father's education that were much more similar to each other than were the combined-data correlations. For example, the combined-data correlation with income was .211 for SAT verbal score, considerably larger than the combined-data correlation of .105 for class rank. By comparison, the average within-high-school correlation with income was .120 for SAT verbal score, slightly *smaller* than the correlation of .129 for class rank.

Table 5

Correlation Results for White Students (N = 62,717) From Zwick and Green (2007)

| Variables | Combined data | Average within-school |
|---------------------------------|---------------|-----------------------|
| Income x SAT Verbal | .211 | .120 |
| Income x SAT Math | .240 | .150 |
| Income x HSGPA | .096 | .101 |
| Income x Class Rank | .105 | .129 |
| Father's Education x SAT Verbal | .320 | .230 |
| Father's Education x SAT Math | .307 | .213 |
| Father's Education x HSGPA | .176 | .178 |
| Father's Education x Class Rank | .190 | .211 |

Note. Adapted from Table 6 of Zwick and Green (2007). HSGPA = high school grade-point average.

Comparable results were obtained by Mattern, Shaw, and Williams (2008), who replicated our analyses with SAT data from 2007. Although additional research is needed, the results to date suggest that, when examined using student-level (within-high-school) correlations, rather than statistics that confound within-school and between-school associations, HSGPA and SAT scores are similarly associated with SES.

Discussion

The unusual nature of HSGPA—its lack of comparability across high schools—is the reason that it produces systematic errors in FGPA predictions and that its association with SES is difficult to assess. It is interesting that despite the variation in grading standards, HSGPA nevertheless has a substantial correlation with FGPA even when data are combined across high schools in the usual fashion (see Table 1). Zwick and Green (2007, p. 42) speculated that the relatively high correlation between college and high school GPA may be attributable to a method effect (Campbell & Fiske, 1959). We can consider admissions test scores, high school GPAs, and college GPAs as alternative measures of academic skill. The two grade-point averages also have a common method, in that they both involve teacher evaluations based on a wide range of student data (see Willingham et al., 2002, p. 19). If a method effect contributes to the HSGPA-FGPA correlation, the within-high-school correlation may be high enough so that, even after reduction due to the noncomparability of HSGPA, the (combined-data) correlation between HSGPA and college GPA still exceeds the SAT-FGPA correlation.

Given this sizable correlation, is it the case that HSGPA is the single best predictor of college performance? A detailed analysis of the interrelationships of grades, test scores, and SES yields a more complex picture. Recent research results show that test scores and high school grades have similar correlations with SES at the student level, a finding that is inconsistent with the claim that HSGPA is less contaminated with SES than are admissions test scores and its relationship with FGPA therefore more authentic. In terms of its predictive value, HSGPA is limited by the variation in grading standards across high schools. This disparity is, in turn, a likely cause of the well-recognized pattern of prediction errors for underrepresented ethnic groups. Zwick and Himelfarb (2011) found that a high school SES index could mitigate the

systematic prediction errors that are typically observed when college grades are regressed on high school grades and admissions test scores.

Would it be advisable to incorporate an index of high school SES or a more broadly defined index of high school quality in admissions decisions? A determination of whether to do so would need to take into account the full array of factors considered in the selection process, the potential impact on the demographics of the entering class, and the mission of the institution. Including an index of this kind in a prediction model for college grades would tend to reduce the predicted grades of underrepresented minority students (assuming these students are more likely than White students to attend poor-quality high schools). If predicted grades were the sole criterion for admission, the result would be to reduce the probability that these students would be admitted. Under this scenario, the incorporation of high school information would work in a direction opposite to that of admissions programs that explicitly or implicitly award preferences to applicants from lower-SES high schools. Among these are the model investigated by Young and Johnson (2004), which involves computation of an SES-adjusted predicted FGPA, and the *percent plans*, which provide for the automatic admission of a fixed percentage of top ranking students from each high school, regardless of its quality (e.g., see Horn & Flores, 2003).

There is, of course, no reason that college admissions need be based on predicted FGPA or even on predicted college performance, defined more broadly. Institutions need not even consider academic talent in admissions, and some do not: Institutions in the *open-door* category require only that candidates submit an application and, in some cases, show proof of high school graduation. In a survey conducted in 2002, 8% of the 957 four-year institutions that responded and 80% of the 663 two-year institutions fell into the open-door category (Breland, Maxey, Gernand, Cumming, & Trapani, 2002, p. 15). For these institutions, the relative predictive value of various academic indicators is unlikely to be of paramount interest.

Most colleges and universities take academic qualifications into account along with other institutional goals, which might include fostering campus diversity, encouraging students with particular talents, or even finding a quarterback for the football team. According to one high school counselor, “those people the admissions dean will want to find include legacies [children of alumni]; leaders for school publications, student government, and other areas of student life; children of influential families; musicians, athletes, public speakers, and others with special

talents; and students from underrepresented ethnic groups and geographical areas” (Mitchell, 1998).

Combining predicted performance with other factors in making admissions decisions is the approach recommended by Tam and Sukhatme (2004), who served as administrators at the University of Illinois at Chicago. After finding that their measure of high-school quality (actually, the average ACT score for the high school) improved the prediction of graduation, they argued that including such a measure in admissions decisions would conserve university resources by increasing the probability that students could successfully complete college work. They noted that “[w]ith the severe constraints on capacity prevalent at [urban public] institutions, admitting students who do not graduate wastes limited educational resources partially subsidized by tax revenue” (p. 13). Tam and Sukhatme concluded that a selection index including students’ ACT scores, their class rank in high school, and their high school’s average ACT score “should be used, together with other considerations, such as diversity, to make better college admission decisions” (p. 12).

Apart from the question of its predictive value, the use of HSGPA as an admissions criterion has several potential advantages: As noted by Geiser and Santelices (2007), focusing on HSGPA in admissions may help to increase ethnic diversity (a claim that is consistent with the results of Table 2). Furthermore, it can serve an important signaling function by communicating to students and parents the message that academic performance in high school is important. Finally, an emphasis on HSGPA can provide an incentive to high school teachers and administrators to improve their academic programs. From this perspective, further research on the advisability of incorporating a high school quality index in the admissions process could have the beneficial effect of focusing attention on the crucial role of high schools in preparing students for college.

References

- Agronow, S., & Studley, R. (2007). *Prediction of college GPA from new SAT test scores: A first look*. Retrieved from <http://www.cair.org/conferences/CAIR2007/pres/Agronow.pdf>
- Atkinson, R. C., & Geiser, S. (2009). Reflections on a century of college admissions tests. *Educational Researcher*, 38, 665–676.
- Bowen, W. G., & Bok, D. (1998). *The shape of the river: Long-term consequences of considering race in college and university admissions*. Princeton, NJ: Princeton University Press.
- Breland, H., Maxey, J., Gernand, R., Cumming, T., & Trapani, C. (2001). *A report of a national survey of undergraduate admission policies, practices, and procedures*. Retrieved from <http://www3.airweb.org/images/trendsreport.pdf>
- Burton, N. W., & Ramist, L. (2001). *Predicting success in college: SAT studies of classes graduating since 1980* (Research Report No. 2001-2). New York, NY: College Entrance Examination Board.
- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56, 81–105.
- Cleary, T. A. (1968). Test bias: Prediction of Negro and White students in integrated colleges. *Journal of Educational Measurement*, 5(2), 115–124.
- Geiser, S., & Santelices, M. V. (2007). *Validity of high-school grades in predicting student success beyond the freshman year* (Research and Occasional Paper Series No. CSHE.6.07). University of California, Berkeley, Center for Studies in Higher Education.
- Geiser, S. (with Studley, R. E.). (2004). UC and the SAT: Predictive validity and differential impact of the SAT I and SAT II at the University of California. In R. Zwick (Ed.), *Rethinking the SAT: The future of standardized testing in university admissions* (pp. 125–153.) New York, NY: Routledge Falmer.
- Horn, C. L., & Flores, S. M. (2003). *Executive summary: Percent plans in college admissions. A comparative analysis of three states' experiences*. Retrieved from <http://civilrightsproject.ucla.edu/research/college-access/admissions/percent-plans-in-college-admissions-a-comparative-analysis-of-three-states2019-experiences/horn-percent-plans-2003.pdf>

- Jencks, C., & Phillips, M. (Eds.). (1998). *The Black-White test score gap*. Washington, DC: Brookings Institution Press.
- 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* (College Board Research Report No. 2008-5). New York, NY: The College Board.
- Linn, R. L. (1983). Predictive bias as an artifact of selection procedures. In H. Wainer & S. Messick (Eds.), *Principals of modern psychological measurement: A festschrift for Frederic M. Lord* (pp. 27–40). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Mattern, K. D., Patterson, B. F., Shaw, E. J., Kobrin, J. L., & Barbuti, S. M. (2008). *Differential validity and prediction of the SAT* (College Board Research Report No. 2008-4). New York, NY: The College Board.
- Mattern, K. D., Shaw, E. J., & Williams, F. E. (2008). *Examining the relationship between the SAT, high school measures of academic performance, and socioeconomic status: Turning our attention to the unit of analysis* (College Board Research Note No. RN-36). New York, NY: The College Board.
- Mitchell, J. S. (1998, May 27). SAT's don't get you in. *Education Week*, 33–34.
- Raudenbush, S. W., & Bryk, A. S. (2001). *Hierarchical linear models: Applications and data analysis methods applications* (2nd ed.). Thousand Oaks, CA: Sage.
- Sackett, P. R., Kuncel, N. R., Arneson, J. J., Cooper, S. R., & Waters, S. D. (2009). Does socioeconomic status explain the relationship between admissions tests and post-secondary academic performance? *Psychological Bulletin*, 135, 1–22.
- Sarbin, T. R. (1943). A contribution to the study of actuarial and individual methods of prediction. *The American Journal of Sociology*, 48, 593–602.
- Soares, J. A. (2012). Introduction. In J. A. Soares (Ed.), *SAT wars: The case for test-optional admissions* (pp. 1–9). New York, NY: Teachers College Press.
- Tam, M.-Y. S., & Sukhatme, U. (2004). How to make better college admission decisions: Considering high school quality and other factors. *The Journal of College Admission*, 183, 12–16.
- Willingham, W. W., Pollack, J. M., & Lewis, C. (2002). Grades and test scores: Accounting for observed differences. *Journal of Educational Measurement*, 39, 1–37.

- Young, J. W. (2004). Differential validity and prediction: Race and sex differences in college admissions testing. In R. Zwick (Ed.), *Rethinking the SAT: The future of standardized testing in university admissions* (pp. 289–301). New York, NY: RoutledgeFalmer.
- Young, J. W., & Johnson, P. M. (2004). The impact of an SES-based model on a college's undergraduate admissions outcomes. *Research in Higher Education, 45*, 777–797.
- Zwick, R. (2002). *Fair game? The use of standardized admissions tests in higher education*. New York, NY: RoutledgeFalmer.
- Zwick, R., & Green, J. G. (2007). New perspectives on the correlation of SAT scores, high school grades, and socioeconomic factors. *Journal of Educational Measurement, 44*, 23–45.
- Zwick, R., & Himelfarb, I. (2011). The effect of high school socioeconomic status on the predictive validity of SAT scores and high school grade-point average. *Journal of Educational Measurement, 48*, 101–121.
- Zwick, R., & Schlemer, L. (2004). SAT validity for linguistic minorities at the University of California, Santa Barbara. *Educational Measurement: Issues and Practice, 25*, 6–16.
- Zwick, R., & Sklar, J. (2005). Predicting college grades and degree completion using high school grades and SAT Scores: The role of student ethnicity and first language. *American Educational Research Journal, 42*, 439–464.

Notes

¹ Sackett et al. (2009) concluded, based on a meta-analysis, that the association between SAT scores and FGPA was virtually undiminished when SES was partialled out. Atkinson and Geiser (2009) criticized the study for failing to incorporate the effects of HSGPA.

² For gender groups, a different pattern holds: Overprediction for men and the corresponding underprediction for women are typically slightly greater in magnitude for a model that includes HSGPA and SAT scores as predictors than for a model that includes HSGPA only. However, average prediction errors tend to be much smaller for gender groups than for ethnic groups. See Zwick and Himelfarb (2011) for further discussion.

³ The small average prediction errors for White students are to be expected because of the high proportion of White students: In 22 of the 34 schools, it exceeds 70%. Therefore the common regression equation in most colleges is presumably similar to the equation that would have been obtained by including only White students in the regression equation.

⁴ A model with SAT only as a predictor yielded prediction errors intermediate between those of Models 1 and 2.