

On the value of standardized admissions tests

In recent months, Wake Forest University has initiated a timely investigation of the college admissions process. Rather than “doing business as usual,” the university has decided to explore alternative patterns of admission tools to determine which methods are most efficient and fair in identifying students we wish to have at Wake Forest. This exploratory process is an important and wise decision, with its aim to increase diversity at the university.

When assessing academic ability, the traditional approach has been to require a high school transcript and scores from a standardized admissions test, most commonly the SAT. As one part of the recent WFU decision, submitting SAT I (a general reasoning test) scores has been changed from being required to being voluntary. The intention of this decision was to diversify the group of students who apply to Wake Forest and to investigate different methods of identifying qualified applicants.

In early fall of 2008, an open letter to the WFU community suggested that a good reason for making the SAT optional was that the test is not predictive of college performance, and therefore should not be a required part of the college admissions process. The letter also implied that the SAT is significantly unfair or biased against some groups of applicants and that it therefore presents difficulties for addressing diversity in the admissions process.

The previous open letter and similar communications have motivated those of us who teach and publish in the area of standardized psychological testing to respond. There are indeed reasons to examine critically the use of the SAT in the admissions process; however, we believe that a lack of predictive ability or the existence of biased predictive validity are *not* valid reasons.

Below, we present data from the University of California (UC) report (Geiser & Studley, 2001), which was cited heavily in the previous WFU open letter. As we demonstrate, the UC Report provides clear evidence that the SAT is indeed a valid predictor of academic success, and that this is true for all groups of students studied to date. We then discuss the SAT’s implications for diversity – implications that can emerge even though the SAT is a validly predictive test. Finally, we discuss criteria that, according to the study of psychological measurement (called psychometrics), should be applied to all methods of evaluating applicants to Wake Forest.

Does the SAT I have value for predicting college performance of applicants?

The question of prediction is traditionally studied in terms of how much the ranking on a set of test scores match the ranking of college grade point average (GPA). The scores might be from the SAT I (a “critical reasoning” test), from high school GPA, or from the SAT II (nationally standardized achievement tests for specific subjects).

The degree of match between test scores and college GPA is indicated by the size of a statistic called a correlation, which ranges, for these purposes, from 0 (no match in rankings) to 1 (a perfect match in rankings). The table below indicates the correlations between various sources of information and college GPA. These correlations are taken from the University of California’s (UC) report mentioned earlier. Information is included for the SAT II because, while WFU does not accept the SAT II, a growing number of colleges do and there is much relevant data available in the UC study.

	Correlation with <u>College GPA</u>
High school GPA	.39
SAT I	.36
SAT II	.40

These correlations show that the SAT and high school grades are very similar in their ability to predict college GPA¹. This is true even though one is based on four years of performance and one is based on a single three-hour experience. Similar findings are reported by many other studies, including one based on data from Wake Forest University (Lawlor, Richman, & Richman, 1997). More generally, Sternberg and the Rainbow Project Collaborators (2006) state that “A recent meta-analysis of the predictive validity of the SAT, encompassing roughly 3,000 studies and more than one million students, suggested that the SAT is a valid predictor of early-college academic performance (as measured by first-year grade point average [GPA])..... The validity coefficients for later-college performance were somewhat lower but still substantial” (p. 322), and they conclude that the SAT “has had substantial success in predicting college performance” (p. 344). Similarly, Sackett et al. (2008) state that “Based on these large-scale studies and meta-analyses, these population correlations suggest that these tests do not predict badly; in contrast, these validity coefficients are quite strong” (p 219)

Thus, an appropriate answer to the question “does the SAT have value for predicting the college academic performance of applicants?” is clearly yes. If one accepts that high school GPA is a valid predictor of college GPA, then one should believe that SAT is a valid predictor of college GPA.

Is it useful to use both the SAT I and high school GPA in predicting college performance?

A second and related issue is whether the combination of SAT *and* high school GPA is better than high school GPA alone. That is, does the SAT provide any useful information,

over and above the information provided by high school GPA? Why require more credentials than you actually need to predict college performance?

To determine whether a predictor provides useful information over and above another predictor, the standard statistical approach is to use “multiple regression” analysis. In the UC Report, the relevant analysis is a two step-process. This analysis first evaluates the predictive power of one predictor (e.g., high school GPA) for college GPA. The amount of predictability can be read as a percent ranging from 0% (no predictability) to 100% (total predictability). The analysis then adds a second predictor (e.g., SAT) to the process, to see if doing so improves the prediction of college GPA. If the second predictor adds no useful information beyond the first, then there should be no change in the “percent of predictability” value from the first step to the second. Results are below.

Regression Results	
	<u>“Percent of predictability”</u>
High School GPA	15.4%
High School GPA and SAT I	20.8%
High School GPA and SAT II	22.2%

Results show that high school GPA predicts about 15% of the variance in college GPA. Although this value may seem small, it is actually quite impressive, considering the complex nature of college performance and the many factors affecting it. Next, the SAT I is added to the analysis along with high school GPA. Results show that, when combined, the two pieces of information account for nearly 21% of the variance in college GPA. A similar analysis shows that, when SAT II is added to high school GPA, the two pieces of information account for approximately 22% of the variance in college GPA. Thus, the SAT tests significantly increase predictability over and above that obtained from high school GPA by itself. This implies that admissions decisions based upon the SAT and high school GPA would be more accurate in predicting college GPA than decisions made without the SAT.

Again, similar results have been demonstrated among students from Wake Forest. The Lawlor et al. (1997) study reported multiple regression analyses in which SAT I was a statistically-significant predictor of Wake Forest GPA even when controlling for high school GPA. In fact, the SAT appeared to be a slightly stronger predictor than did high school GPA in this analysis (note that the authors did not report results for African American students, arguing that there were too few African American participants in the sample to support this more statistically-demanding analysis).

Critics of the SAT often argue that high school GPA is a better predictor than the SAT, but the results presented above indicate that such criticisms are misleading. As has been shown, the UC Report indicates that the difference is actually minimal. In addition, admissions decisions are not based upon a single piece of information – thus, there is little value to debating which is the best single predictor. The UC Report clearly indicates that the SAT can provide usefully valid information, above and beyond high school GPA.

Does the SAT I predict college performance for all socio-economic and racial/ethnicity groups?

An important issue in applied testing is whether a standardized test is equally valid for different groups of people. Of greatest concern in the present discussion is whether the SAT is unbiased and predictive for people from different socio-economic status (SES) and for people of different racial/ethnic background. Fortunately, relevant information is available in the UC Report. Because the UC Report examines the two issues (i.e., SES and race/ethnicity) through different analytic approaches, we discuss them separately.

Socioeconomic status

The earlier open letter to Wake Forest pointed out that the SAT, unlike high school GPA, is correlated with socio-economic status (SES). According to a UC analysis, the correlation of family income with SAT I (Verbal) is .32, with SAT I (Math) is .24, and with high school GPA is .04. This merits serious attention, and it will be discussed later. However, a first question should be whether this correlation makes a difference in its ability to predict academic success? That is, even if such a correlation exists, does it mean that the SAT is not predictive of college performance for persons of all socioeconomic statuses?

To study the relationship between the SAT and college GPA across socio-economic groups, multiple regression is again an appropriate analysis. As mentioned earlier, to determine whether two predictors provide usefully-different information, the standard approach is to put them into a multiple regression equation. For the analyses below, SAT scores were put into equations that also included information about socioeconomic status, with college GPA as the outcome to be predicted. This approach reveals the predictive value of a single predictor once you eliminate any overlap it has with other predictors in the analysis. The numbers presented below are taken from data on the UC study website. In these analyses, the value next to a predictor below indicates the “unique” amount of predictive power (here interpreted somewhat like a correlation, rather than as a percentage) between that predictor and college GPA, above and beyond the predictive power provided by the other predictors in the equation. If the SAT’s predictive power is affected by its association with SES, then its unique predictive value would be very close to zero. However, if the SAT’s predictive power is *not* affected by any association with socio-economic status, then its value should remain well above zero (and close to the values in the first table, above).

Relevant regression analyses from the University of California show that, even when socio-economic status is taken into account, the predictive power of the SAT remains significantly above zero. As summarized below, when SES is controlled in the form of parental education and family income, the predictive power of the SAT-I remains essentially unchanged, dropping only slightly from $\beta = .36$ to $\beta = .345$ (p. 177, 184, UCOP). The SAT-II shows a similar pattern of results – correlating with college GPA at $\beta = .40$ (p. 185), which is reduced only slightly (to $\beta = .375$, p. 178) when SES is taken into account. Thus, the link between the SAT and college GPA remains solid, even when adjusted for any differences in students’ socio-economic statuses.² Recently, this finding

has been reiterated more broadly in a major scholarly journal in the field of Psychology (Sackett et al., 2009).

Socio-economic status and SAT I as predictors of college GPA		Socio-economic status and SAT II as predictors of college GPA	
	Unique predictive association with College GPA		Unique predictive association with College GPA
Family income	.01	Family income	.02
Parental education	.04	Parental education	.06
SAT I	.345	SAT II	.375

Racial/ethnic group

Another extremely important issue is whether the SAT is predictive of college GPA for students from different racial/ethnic groups. If differences exist across groups on test performance, then the test may not be equally predictive for all groups. The UC Report again addresses this point. The correlations given below (derived from values in Table 7 of the report) are between college GPA and high school GPA, SAT I, and SAT II for each group.

	<u>Correlation with College GPA</u>		
	<u>High school GPA</u>	<u>SAT I</u>	<u>SAT II</u>
African American	.31	.32	.35
American Indian	.30	.29	.33
Asian American	.40	.35	.41
Chicano/Latino	.35	.33	.32
White	.39	.32	.37

As can be seen from these correlations, the SAT tests predict performance for all racial/ethnic groups of applicants at similar levels. The values vary to some degree, but generally they are highly stable across all groups, and they are very similar to those of high school GPA.

Furthermore, for each of the five racial/ethnic groups, the UC Report demonstrates that the combination of SAT *and* high school GPA is better than high school GPA alone. Returning to the logic of “percent of predictability” discussed earlier, the UC Report examines presents a series of analyses revealing the predictive benefit of combining SAT scores with high school GPA. First, the UC Report presents the degree to which high school GPA alone predicts college GPA for each racial/ethnic group. As shown in the first column of numbers in the Table below (adapted from Table 7 in the UC Report), these values range from 8.8% to 15.9% for the different groups. Next, the UC Report presents the degree to which the combination of SAT and high school GPA predicts college GPA for each racial/ethnic group. As shown in the second column of numbers, the combination of SAT I and high school GPA produces predictability ranging from

12.5% to nearly 21% for different racial/ethnic groups. More specifically, the increase in predictive power obtained with the SAT I ranges from a low of 3.5% (for White students, $19.1 - 15.6 = 3.5$) to a high of 5.5% (for African American students, $15 - 9.5 = 5.5$). As shown in the third column of numbers below, the combination of SAT II and high school GPA produces even greater increases in predictive power.

	<u>“Percent of Predictability”</u>		
	<u>High school GPA alone</u>	<u>SAT I and HSGPA</u>	<u>SAT II and HSGPA</u>
African American	9.5%	15.0%	16.4%
American Indian	8.8%	12.5%	13.9%
Asian American	15.9%	20.8%	23.4%
Chicano/Latino	12.0%	17.3%	16.4%
White	15.6%	19.1%	20.9%

Issues with using the SAT in admissions: Implications for diversity

If indeed the SAT is useful in predicting college performance in general and for a range of groups, then why is there concern regarding its use? There are a number of social concerns, such as the intense pressure on high school seniors to perform on this test. However, the most serious concern is that using the test may have negative implications for certain groups. For example, there are data showing consistent differences in average scores across different SES and racial/ethnic groups. In the world of applied psychological testing, this is called the adverse impact problem. When making selection decisions, a test might produce differences in the selection rate from different social groups, even if that test itself is clearly valid, predictive, and psychometrically unbiased.

The earlier open letter to Wake Forest implied that the different selection rate from various groups is due to clear bias or unfairness in the SAT I. It should be noted that this issue is quite controversial and has been the subject of a long-standing debate and analysis in the testing area. One common, though not well-supported, contention is that the test contains materials on which only certain groups are likely to succeed. The possibility of such overt bias may have been true in earlier decades, but for some time the College Board has been using a variety of techniques (statistical analyses and panels of minority members) to reduce such bias.

The fact that a difference does exist between average scores on the SAT by socio-economic status and racial/ethnic group is not prima facie evidence of bias. Indeed, a test may be highly predictive of college performance, have no overt bias, but still show an average difference among applicant groups because of other factors. To name a few such factors, there may be differences in the educational opportunities for different groups; there may be differential opportunities to use coaching services to prepare for the test; and there may be a differential “stereotype threat” of the test to certain groups if they perceive the test as biased.

In other research on this point, the evidence is very mixed on the issue of SAT scores, SES, race/ethnicity, and prediction of academic performance. For example, Rothstein (2004) conducted analyses indicating some minimizing effects of SES on the SAT's predictive power (though not nullifying). However, Zwick and Green (2007) conducted analyses indicating that: a) high school GPA might be more strongly linked to SES than is often reported, and b) SAT might be less strongly linked to SES than is often reported. Such analyses suggest that the SAT is even less potentially problematic than is sometimes assumed (and high school GPA more problematic). Moreover, other research indicates that, SAT scores may over-predict for African American students - despite Dr. Freedle's claims to the contrary (cited in a previous open letter to Wake Forest), it is not clear that the content of the test is culturally biased or that supposed content biases create differential test scores. Finally, research by Lawlor et al. (1997) indicates a racial difference on SAT scores, but no racial differences on college GPA or on high school GPA. Such findings need to be replicated, but Sackett et al. (2008) point out that such findings could result from other factors affecting college GPA in the opposite direction of SAT scores. Clearly, the connection between SES, race/ethnicity, SAT, HS GPA, and college GPA is a very complex issue that merits continued discussion. However, the balance of research seems to indicate that it is inaccurate to conclude that SAT scores are biased or hopelessly compromised by SES or race/ethnicity. However, we again acknowledge that the existence of any links between the SAT and SES or race/ethnicity should motivate serious, well-informed consideration – even though there is not consistent evidence that such links arise from invalidity or bias in the SAT itself.

These factors present serious societal issues, but should not necessarily add up to a general condemnation of the test itself, nor do they preclude using the test in a manner that is cognizant of the potential problem of adverse impact. Our inspection of relevant data reveals that the SAT is, in fact, predictive of college performance and as such could be useful. It is notable that the University of California system, even though it produced the data cited in the current and previous letters, continues to use the SAT in its admission process. Given that the SAT's predictive power is not clearly affected by its association with socio-economic status, and given that its predictive power is fairly stable across diverse racial/ethnic groups, there may be some method of using scores from a nationally standardized test that avoids the problem of adverse impact and produces an even-handed identification of qualified applicants.

Desirable aspects of predictors of college performance

WFU has taken the initiative in studying how to handle admission procedures. Indeed, one contribution WFU could make to the national debate on college admissions is to study intensively how to identify and measure the personal qualities we are looking for in a desirable applicant. We applaud such efforts and simply make this final note. Any system of admissions should be evaluated according to the criteria developed in psychometrics to ensure quality in the admissions process. In past research and discussion, heavy scrutiny has focused on how standardized tests stand up to these criteria. However, all methods used for admission should be evaluated according to these

criteria – whether high school grades, interviews, letters of recommendation, or personal essays. The important criteria are:

1) Reliability

In its simplest form, this means that if you have a test score, interview rating, grade or essay score for an applicant, you should be confident that the score or rating would be roughly the same if the applicant were evaluated on another day, by another interviewer, or on another version of the test.

2) Validity

In a simplified form, this means that the score or rating should accurately reflect the actual quality of the applicant. The difference in a high or low score must accurately portray a real difference between applicants. For a method to have validity, one must be sure that both the predictor (such as a test) and the criterion (the actual quality being predicted) are both measured with reliability.

3) Wide range of scores

A problem emerges when a method produces only a limited range of scores or ratings for a group of applicants. This is one of the criticisms of “grade inflation.” If everyone earns an “A,” how can we tell levels of ability apart? Whatever methods we choose should be precise enough to make discriminations among applicants so that we can truly select the students we wish to form the new class at Wake Forest.

4) Lack of bias

The possibility that a test, an interview process, or high school grade point average is biased or unfair for a group of applicants is a serious issue demanding understanding and attention. Any piece of information to be considered in the admissions process should be relatively free of strong bias. As we have discussed before, bias and adverse impact are complex, separable issues. A serious evaluation of these issues must be informed by a clear understanding of their meaning, their causes, and their implications.

Summary

We fully support a thorough and well-informed evaluation of the admissions process at Wake Forest. Toward that end, the purpose of this document is to provide insight into the value of standardized testing in this process. Indeed, we are optimistic that a fair, thorough, and psychometrically-informed analysis of the entire admissions process can help the university attract and identify a diverse and talented group of students to join the Wake Forest community.

R. Michael Furr and Cecilia H. Solano,
Associate Professors, Department of Psychology, Wake Forest University

Notes

1 These results contradict assertions made to the Wake Forest community in the open letter dated Sept 24, 2008, such as: “[Geiser and Studley]... found that high school grades (HSGPA) and subject specific tests, such as the SAT II, strongly predicted college grades, but not the SAT.” Such statements seem to be based on analyses that do not clearly support them. Specifically, such statements appear to be based on the last 2 rows in Table 1 of the UC Report – Row 7 and the “SAT I increment” row (e.g., the previous open letter stated that “In two of the four years they researched, the SAT was not a statistically significant variable in their multivariate regression models for predicting grades”). Unfortunately, the analyses reported in these two rows do not address the basic predictive power of the SAT-I. Rather, they reflect the predictive power of the SAT-I, after controlling for high school grades AND after controlling for the SAT-II. Thus, the appropriate conclusion from such analyses is that it is unnecessary to use both the SAT-I and the SAT-II together, but this is irrelevant to the basic predictive power of either test in its own right.

2. These results again contradict explicit and implicit assertions made in the previous open letter to Wake Forest dated Sept 24, 2008 – assertions such as “SES differences do not reduce HSGPA or subject test, SATII, effects, but they do reduce SATI effects to an extent not worth noticing.” Such assertions (seemingly about the effect of SES on the basic, absolute predictive power of the SAT-I) were based upon analyses that do not speak to the basic predictive power of the SAT tests.

Specifically, the previous open letter to Wake Forest cited and reprinted Table 6 from the UC Report to support such assertions. Each column in Table 6 presents results from a separate multiple regression equation in which three or more variables were used to predict college GPA. Multiple regression (i.e., a regression equation including more than one predictor variable) reveals the “unique” predictive power of a variable – the predictive power of a variable over and above the other variables in the equation. The first column presents regression weights for an equation that includes HSGPA, SAT-I, and SAT-II. This column reveals that, after controlling for the fact that the SAT-I is correlated with both HSGPA and SAT-II, the SAT-I it has relatively low unique predictive power ($\beta = .07$, for this table, values close to zero indicate no predictive power). Furthermore, the second column reveals that, after controlling for the fact that the SAT-I is correlated with HSGPA, controlling for the fact that it is correlated with SAT-II, and controlling for two variables related to SES, the SAT-I has minimal unique predictive power (i.e., $\beta = .02$).

Despite the previous open letter’s reference to these analyses, they in fact provide no information about the degree to which the SAT’s basic predictive power is compromised by SES. *Rather, they tell us that: a) if we already have HSGPA and SAT-II information, then the SAT-I isn’t very useful, and b) if we have SAT-II, HSGPA, and SES information, then the SAT-I is somewhat less useful.* In actuality, the apparently small predictive values for the SAT-I reflect the fact that SAT-I scores are extremely highly correlated with SAT-II scores. Close inspection of the supplemental analyses available on the UCOP web site reveals that the two tests are highly correlated at $r = .84$, signifying tremendous overlap between the two tests – that is, the two tests provide very similar information. Furthermore, as described above and as shown in Table 6, both SAT-I and SAT-II were included in each of the two regression models. In a multiple regression equation, the apparent predictive power of one variable is limited by its overlap with *all other predictors* in the equation. It is the extreme overlap between SAT-I and SAT-II that produces the very low values for the SAT-I in Table 6. Therefore, the values in Table 6 tell us nothing about the degree to which SAT-I’s basic predictive power is compromised by SES. Rather, they tell us that, if we already have SAT-II scores, then we don’t need SAT-I scores. It should be noted that, at present, Wake Forest does not require applicants to submit SAT-II scores.

We are not the first to point out that the values in Table 6 of the UC Report provide no information about the effect of SES on the SAT’s basic predictive power. This point has been made before (e.g., Johnson, 2004; Sackett et al., 2008), and it follows directly from the basic principles of multiple regression analysis.

References

All of these are available online or through the Wake Forest University libraries

- Geiser, S., & Studley, R. (2001, October). *UC and the SAT: Predictive Validity and Differential Impact of the SAT I and SAT II at the University of California*. Oakland, CA: University of California Office of the President. Available at <http://www.ucop.edu/sas/research/researchandplanning/>
- Johnson, J. W. (2004). Not all affirmative action rewards merit: Comment. *American Psychologist*, 59, 123-124.
- Lawlor, S., Richman, S. & Richman, C.L. (1997). The validity of using the SAT as a criterion for Black and White students' admission to college. *College Student Journal*, 31, 507-15.
- Rothstein, J. (2004). College performance predictions and the SAT.” *Journal of Econometrics*, 121, 297-317.
- Sackett, P. R., Borneman, M. J., & Connelly, B. S. (2008). High-stakes testing in higher education and employment: Appraising the evidence for validity and fairness. *American Psychologist*, 63, 215-227.
- 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.
- Sternberg, R. J., & The Rainbow Project Collaborators (2006). The Rainbow Project: Enhancing the SAT through assessments of analytical, practical and creative skills. *Intelligence*, 34, 321-350.
- Zwick, R. & Green, J.G. New perspectives on the correlation of SAT scores, high school grades, and socioeconomic factors. *Journal of Educational Measurement*, 44, 23-45. 2007.
- UCOP (n.d.). Supporting Documents for UC and the SAT: Predictive Validity and Differential Impact of the SAT I and SAT II at the University of California: Regressions.rft. Downloaded October, 24, 2008 from <http://www.ucop.edu/sas/research/researchandplanning/supporting.htm>.