ARTICLE IN PRESS

Available online at www.sciencedirect.com

SCIENCE DIRECT

Journal of Econometrics ■■■ (■■■■) ■■■–■■■

JOURNAL OF Econometrics

www.elsevier.com/locate/econbase

College performance predictions and the SAT

Jesse M. Rothstein*

Department of Economics & Woodrow Wilson School of Public and International Affairs, Princeton University, Industrial Relations Section, Firestone Library, Princeton, NJ 08544, USA

Abstract

The methods used in most SAT validity studies cannot be justified by any sample selection assumptions and are uninformative about the source of the SAT's predictive power. A new omitted variables estimator is proposed; plausibly consistent estimates of the SAT's contribution to predictions of University of California freshman grade point averages are about 20% smaller than the usual methods imply. Moreover, much of the SAT's predictive power is found to derive from its correlation with high school demographic characteristics: The orthogonal portion of SAT scores is notably less predictive of future performance than is the unadjusted score.

JEL classification: C24; C53; I20; J78

Keywords: SAT validity; College admissions; Sample selection; Omitted variables

1. Introduction

Several prominent colleges have recently de-emphasized the SAT entrance exam in admissions. The University of California (UC), perhaps in reaction to a mid-1990s prohibition on race-based affirmative action (U.S. Commission on Civil Rights, 2000), has been particularly active. In 2001, it implemented the "Four Percent Plan," whereby students at the top ranks of their high school classes are guaranteed admission regardless of SAT scores, and in 2002 it adopted a "comprehensive review" plan that will inevitably reduce the importance of SAT scores in campus admissions. Responding to UC President Richard Atkinson's (2001) proposal to eliminate the UC's SAT requirement, the College Board recently announced substantial changes to the SAT exam.

Critics argue that SAT scores measure academic preparedness and that any dilution of the SAT's role will lead to a less qualified entering class (Barro, 2001; McWhorter,

* Tel.: +1-609-258-4045; fax: +1-609-258-2907.
E-mail address: jrothst@princeton.EDU (J.M. Rothstein).

0304-4076/$ - see front matter
doi:10.1016/j.jeconom.2003.10.003

2001). In this view, the acknowledged correlation between SAT scores and student socioeconomic status is an unfortunate side effect of educational inequality: Students from disadvantaged backgrounds are simply not as well prepared to succeed in college, and the SAT should be credited, not blamed, for measuring this shortfall (Camara, 2001).

Granting that collegiate academic success is the appropriate maximand of the admissions process, the debate hinges on an empirical question: How much information does the SAT provide about a student's future performance in college? To evaluate the UC's new Four Percent Plan, for example, one wants to know whether the high GPA, low SAT students it admits will perform better or worse than the lower GPA, higher SAT students they displace.

A long "validity" literature attempts to answer this question by estimation of prediction models for collegiate outcomes like freshman grade point average, FGPA (see, e.g., Bridgeman et al., 2000; Camara and Echternacht, 2000; Stricker, 1991; Willingham et al., 1990). Unfortunately, the validity literature is not especially informative about the effects of changes in admissions policies. Two methodological shortcomings stand out. First, prediction models are estimated on highly selected samples, with theoretically indeterminate effects on the estimated SAT contribution. It is demonstrated below that no selection assumptions can justify the approach typically taken. Second, validity studies typically fail to take account of other variables that predict college performance, and are therefore uninformative about the source of the SAT's predictive power. The usual methods would assign predictive power to any candidate variable that happened to correlate with students' demographic and socioeconomic characteristics, and thus necessarily overstate the SAT's importance.[1]

This paper attempts to correct these shortcomings in estimating the SAT's validity for predictions of FGPA at the University of California. I take advantage of a helpful feature of the UC's admissions process: Many applicants are guaranteed admission to the UC on the basis of SAT scores and high school grades alone. An analysis sample composed of these students is free of an important source of sample selection that often biases prediction model coefficients. Under the (admittedly strong) assumption that individual campus admissions and student matriculation decisions are ignorable, OLS prediction coefficients are consistent as long as both the SAT score and the high school grade point average are included as independent variables. A new omitted variables estimator is used to extend these results to the restricted models needed for conventional validity measures. The resulting estimate of the SAT's predictive contribution is about 20% lower than that found by traditional methods. Several alternative specifications designed to address additional forms of sample selection indicate that this may slightly *overstate* the true, selection-free contribution.

[1] Researchers often estimate prediction models within racial groups (Young, 2001). Single-group validities, however, do not permit evaluation of policies that operate both across- and within-groups. Willingham and Breland (1982) and Willingham (1985) use an approach similar to that taken here, but do not focus on implications for the SAT's role. Bowen and Bok (1998) and Crouse and Trusheim (1988) take other approaches.

I also investigate the sensitivity of the SAT's apparent predictive power to the inclusion of student background characteristics as FGPA predictors. Variables describing the demographic composition of students' high schools are found to be strong predictors of both SAT scores and FGPAs. Moreover, the latter are substantially more strongly related to the portion of students' SAT scores that can be predicted from background characteristics than to the part that cannot. The SAT apparently serves in part as a proxy for omitted background variables in sparse models, which therefore overstate its incremental informational content. My preferred estimate of the SAT's predictive contribution in the UC sample is 60% lower than would be indicated by traditional methods. The richer models do not extend directly to admissions policy, as they imply a politically infeasible—and possibly undesirable—admissions rule. Nevertheless, the results here raise questions about whether FGPA-maximizing admissions policies can be seen as meritocratic.

In order to focus on the questions at hand, several topics emphasized in the validity literature are neglected. Most importantly, I assume an underlying single index of student preparedness, an increment to which produces the same change in expected FGPA at all campuses and in all majors. Most validity studies make weaker assumptions, permitting prediction coefficients to vary without restriction at different colleges (Breland, 1979; Young, 1993, reviews other approaches). Doing so makes sample selection problems intractable. It is also inconsistent with admissions practice: The UC uses a single index of SAT scores and high school grades for determining eligibility to any its eight campuses. [2] As the UC's systemwide eligibility rules are the most relevant to many proposed policies, I maintain the single-index assumption and estimate prediction models for a pooled sample of UC students.

2. The validity model

The admissions office's (stylized) problem is to identify a subset of the applicant pool most likely to be academically successful. The office's assessment of student i may be written as

$$\mathrm{E}[y_i^*|X_i, S_i] = \alpha + X_i\beta + S_i\gamma, \tag{1}$$

where S_i is the student's SAT score, X_i is a vector of other admissions variables, and y_i^* is a measure of the student's preparedness for college. [3] Given observations on a realization of y^*, y, for a random sample of students, best linear predictor coefficients α, β, and γ can be estimated by OLS regression of y on X and S. Predictive accuracy is measured by the regression goodness-of-fit.

In SAT validity studies, X typically consists only of a single variable, the high school grade point average (HSGPA) or class rank, while y is the freshman grade point average (FGPA). Studies generally measure the SAT's importance by two statistics:

[2] Hernandez (1997) reports that the Ivy League colleges also share a common admissions index.

[3] I assume throughout that the conditional expectation is linear and additive. A variety of semiparametric models, not reported here, offer no evidence against this assumption, and indeed indicate a remarkably linear relationship in the scaled SAT score.

the fit of a restricted model that forces $\beta = 0$ and the increment to fit conferred by the unrestricted model (1) over another model that forces $\gamma = 0$. When goodness-of-fit is measured as $R = \sqrt{R^2}$ these statistics are known as the SAT's *raw validity* and *incremental validity*, respectively.

2.1. Restriction of range corrections

The population for whom outcomes are to be predicted is the group of potential applicants, while FGPA is observed only for students who have previously matriculated, likely not a representative group. Sample selection problems are compounded by the focus on goodness-of-fit statistics rather than regression coefficients, as the former may be inconsistent even when the latter are estimated without bias. "Restriction of range" corrections are used to resolve this problem (Camara and Echternacht, 2000; Willingham et al., 1990). These corrections are described here in the context of estimating the population explained share of variance.[4] It is then shown that no assumptions about the sample selection process can justify the typical use of range corrections.

Suppose that the data generating process for y is linear, additive, and homoskedastic:

$$y_i = \alpha + X_i\beta + S_i\gamma + \varepsilon_i, \tag{2}$$

where ε_i is independent of (X_i, S_i) with zero mean and constant variance σ^2. The theoretical explained share of variance is

$$R^2 \equiv 1 - \frac{\operatorname{var}(y|X,S)}{\operatorname{var}(y)} = 1 - \frac{\sigma^2}{(\beta' \;\; \gamma')\Sigma(\beta' \;\; \gamma')' + \sigma^2}, \tag{3}$$

where Σ is the variance–covariance matrix for (X, S).

Now suppose that (2) and (3) are estimated on a sample selected purely on the basis of X and S. With this selection assumption, the OLS estimates $\hat{\beta}, \hat{\gamma}$, and $\hat{\sigma}^2$ are consistent. However, the within-sample $\hat{\Sigma}$ is not consistent for its population value. Corrections for "restriction of range" recover an estimate of R^2 by using a consistent estimate of Σ, drawn from an auxiliary, unselected data source, in place of the within-sample variance–covariance in (3).

2.2. The logical inconsistency of range corrections

The range corrected R^2 is consistent for the population explained share of variance as long as OLS coefficients and residual variance are themselves consistent. As noted, one condition that assures this is selection-on-observables: If sample selection (via admissions and matriculation decisions) is independent of ε_i, consistency is guaranteed. Even in this unlikely situation, however, usual practice cannot be justified. The typical study also estimates restricted versions of (2) in which either β or γ is constrained to be zero. Altering the notation slightly, three regression models must be estimated:

$$\mathrm{E}[y_i|X_i, S_i] = \alpha_1 + X_i\beta_1 + S_i\gamma_1 \tag{4A}$$

[4] Empirical results below also present traditional validities. As we are concerned with consistency rather than finite-sample unbiasedness, the extension from R^2 to R is trivial.

$$E[y_i|X_i] = \alpha_2 + X_i\beta_2 \tag{4B}$$

$$E[y_i|S_i] = \alpha_3 + S_i\gamma_3 \tag{4C}$$

The last of these is required for SAT's raw validity, considered in the absence of other variables; the first and second are needed to estimate SAT's incremental validity. Letting $\tilde{R}_j$ denote the range-corrected R from model j, both $\tilde{R}_C$ and $\Delta \equiv \tilde{R}_A - \tilde{R}_B$ are reported.

Selection on (X, S) is not sufficient for consistency of OLS for (4B) and (4C). Estimation of (4B) requires that sample selection be on X alone; otherwise, selection on S biases the within-sample regression of S on X. The restricted model relies on this regression—rewrite (4B) as $E[y|X] = E[E[y|X, S]|X] = \alpha_1 + X\beta_1 + E[S|X]\gamma_1$—and is inconsistent when S enters the selection rule. A similar argument shows that estimation of (4C) requires that X be irrelevant to selection.

Taken together, OLS coefficients and range-corrected fit statistics are consistent for (4B) and (4C) only when sample selection is independent of (X, S, ε). In this case, however, there is no restriction of range and OLS is consistent for all desired statistics.

Section 4 describes a new omitted variables approach that takes advantage of the population Σ used for range corrections to permit consistent estimation of (4B) and (4C) from the unrestricted model (4A). Even this requires an unusual sample selected only on (X, S). Most admissions rules use indicators of student preparedness—recommendation letters, extracurricular activities, essays—that cannot be controlled for in X, making the selection-on-observables assumption untenable. A crucial part of the University of California admissions process, however, considers only observed variables. The next section describes the construction of an analysis sample in which the assumption is plausible, at least for one important selection margin.

3. Data

I use an unusually large and rich data set extracted from University of California administrative records. The data contain observations on all 22,526 California residents (18,587 with complete records) from the 1993 high school class who applied to, were admitted by, and enrolled as freshmen at any of the eight UC campuses.[5] This cohort predates the recent changes in UC admissions policies.

The data, while rich, impose several limitations on the analysis. Most importantly, the only test score available is the SAT score composite, the sum of separate math and verbal scores.[6] All models estimated here thus impose identical prediction coefficients

[5] Observations from one of the campuses, Santa Cruz, are excluded from all analyses: Grades are optional at UCSC and are infrequently assigned. A ninth campus, San Francisco, enrolls only graduate students.

[6] Since 1994, the SAT is called the SAT I, and is marked on a "recentered" scale. This paper uses "SAT" and "SAT I" interchangeably, with all scores on the pre-1994 scale.

for the two SAT subtests. Moreover, I cannot test whether the results generalize to the subject-specific SAT II exams, though Geiser and Studley's (2001) analysis of similar data suggests that they would not.

The criterion used is the UC-wide FGPA.[7] I assume that a change in preparedness produces the same increment to FGPA at all campuses and in all courses. It may induce a change in campus or course selection, which I allow to affect FGPA through campus and freshman major fixed effects. This is in the spirit of a matching procedure used by Goldman and Widawski (1976) to correct the FGPA for departmental differences in grading standards, though the lack of course-level enrollment information prevents implementation of their full procedure. Campus and major effects are interpreted as criterion adjustments, not as explained variance, so are excluded in calculations of goodness-of-fit statistics. In keeping with this interpretation, they are constrained to remain unchanged in estimates of the restricted models (4B) and (4C). Of course, both campus and major are potentially endogenous. Several specification checks presented in Section 5 suggest that endogeneity of campus and major does not seriously bias prediction coefficients from the pooled sample.

Two auxiliary data sources are used in concert with the UC database. First, data from the College Board, with observations on all California SAT-takers from the 1994 to 1998 high school cohorts, are used to estimate Σ, the population variance–covariance matrix of SAT scores and high school GPAs. Second, school level demographic characteristics are drawn from the California Department of Education's Academic Performance Index (API) database.[8] The API data cover only public schools, and analyses in Section 6 are therefore restricted to graduates from 671 public high schools (81% of the students in the UC database). Range corrections in that section extend results only to the population of public school SAT-takers.

Table 1 reports summary statistics for the population of California SAT-takers, for the UC sample, and for a subsample consisting of "UC eligible students."

3.1. UC admissions processes and eligible subsample construction

The UC system's mandate is to admit the top 12.5 percent of California high school graduates each year. Admissions decisions are made in several stages. First, a central office determines whether each applicant is *eligible* to the UC (that is, whether she is in the top 12.5%).[9] All eligible students are guaranteed admission to at least one campus. Second, each of the campuses to which the student applied decides whether to *admit* her. Campus admissions give preferences to eligible students, and no campus may make more than 6% of its admission offers to ineligible applicants. Because campus admissions are not centrally coordinated, some eligible applicants are admitted to

[7] An earlier version of this paper (Rothstein, 2002) also estimated models for two longer-term measures. Prediction models for the overall college GPA resemble those for the freshman year. The SAT score seems to have less predictive power, however, for graduation rates than for grade-based measures. Wilson (1983) reviews similar results.

[8] The API data were first collected in 1999, and may measure 1993 school characteristics with error.

[9] UC eligibility policies—the Four Percent Plan, for example, but not affirmative action, which was only ever practiced at the individual campus level—are thus about how to define the "top 12.5%."

Table 1
Summary statistics for UC matriculant and SAT-taker samples

	UC sample				CA SAT-takers, 1994–1998	
	All		UC eligible only			
	Mean (A)	*S.D.* (B)	*Mean* (C)	*S.D.* (D)	*Mean* (E)	*S.D.* (F)
Number of observations	18,587		17,346		620,013	
FGPA	2.84	0.64	2.88	0.62		
SAT	1,091	179	1,104	172	904	226
HSGPA[a]	3.75	0.46	3.81	0.39	3.24	0.63
Black	4%		3%		7%	
Hispanic	14%		12%		19%	
Asian	39%		40%		22%	
Female	53%		53%		55%	
% with school data	81%		81%		76%	
Skl: Frac. Black[b]	7%	10%	7%	10%	8%	11%
Skl: Frac. Hispanic	27%	23%	27%	22%	30%	24%
Skl: Frac. Asian	21%	18%	21%	18%	17%	17%
Skl: Frac. lunch	25%	22%	24%	21%	28%	22%
Skl: Avg. parental ed.	14.4	1.3	14.5	1.3	14.2	1.3

[a]HSGPAs in columns A–D are calculated using the UC's weighting rule, which assigns 4 points to an A grade and awards an extra point to grades earned in honors courses.

[b]School variables report means and standard deviations of student-level measures among the public school students for which the data are available.

several campuses while others are rejected from all the campuses to which they have applied. In a third stage, the latter students are offered admission at one of the less selective campuses, frequently Riverside.

The central eligibility determination, unlike campus admissions decisions, is based on published rules that in 1993 consisted primarily of a deterministic function of the HSGPA and composite SAT score (UC Office of the President, 1993). As a result, eligibility can be simulated for each student in the UC database. A subsample is constructed consisting of the 17,346 students (14,102 from public schools with API data) who are judged to have been UC eligible.[10]

For students in the eligible subsample, admission was guaranteed, and sample selection came only from decisions to apply and to accept an admissions offer that may not have been at the preferred campus. If these decisions may be assumed uninformative about unobserved ability—ε_i in (2)—the subsample is selected on observables and permits consistent OLS estimation of (4A).[11] Moreover, even if student enrollment decisions are informative about ability, the subsample is arguably representative

[10] These are 93.3% of the UC sample; slightly more than 6% of the students in the sample are ineligible. This overrepresentation relative to the above-cited 6% limit likely reflects different matriculation rates—yields—among eligible and ineligible admitted students.

[11] Leonard and Jiang (1999) note the utility of the same institutional feature for validity studies.

(again, conditional on X and S) of the population of interest to admissions offices, who presumably care only to accurately predict performance for students who might choose to enroll if admitted. In the next section, I present validity estimates, first using conventional, inconsistent methods and then taking advantage of the UC eligibility rule to obtain estimates that are consistent as long as matriculation, campus, and major choices are ignorable. Section 5 examines this assumption, presenting evidence that student self-selection into the UC and into particular campuses does not substantially bias pooled-model coefficients.

4. Validity estimates: sparse model

Table 2 presents several estimates of the basic validity model described in Section 2, using HSGPA as the only X variable. The first column presents OLS estimates from the full UC sample. Panels A–C report coefficients and range-corrected fit statistics from models (4A) through (4C), respectively. The final rows report the SAT increment to goodness-of-fit, the difference between fit statistics in Panels A and B. The usual validity methods would thus report the SAT's raw validity as 0.490 and its incremental validity as 0.055.

The regression coefficients in Column A are potentially biased by endogenous sample selection. Column B repeats the models using only the UC-eligible subsample. The eligibility determination considers only observables, so coefficients from the unrestricted model in Panel A should not be biased by *eligibility-induced* sample selection.[12] The sample of eligible students produces a 13% increase in the HSGPA coefficient, with a negligible effect on the SAT coefficient. (The differential effect probably reflects the greater importance of HSGPA in the 1993 eligibility determination: Table 1 reveals that ineligible students have low HSGPAs but SAT scores comparable to those of eligible students.)

Panels B and C of Column B present OLS estimates of the restricted models from the eligible subsample. These estimates are inconsistent, as observed independent variables in (4A) act as unobservables in (4B) and (4C). Recall the omitted variables formulation of (4B):

$$\mathrm{E}[y|X] = \mathrm{E}[\mathrm{E}[y|X,S]|X] = \alpha_1 + X\beta_1 + \mathrm{E}[S|X]\gamma_1. \quad (5)$$

Because eligibility rules allow high SAT scores to compensate for low HSGPAs (in a narrow range), $\mathrm{E}[S|X]$ is likely to be substantially different in the sample than in the population. Fig. 1 displays kernel estimates of this conditional expectation in the SAT-taker population, in the full UC sample, and in the eligible subsample. The uptick at the leftmost extreme of the subsample graph reflects the SAT-HSGPA tradeoff inherent in the eligibility rule. The figure legend reports linear regressions of S on X; these differ across samples not only in their vertical positions—which would be absorbed by the intercept in (5)—but also in their slopes.

[12] In the presence of other forms of sample selection, these coefficients may yet be biased. In Section 5, I offer evidence that endogenous matriculation and campus selection do not introduce substantial bias.

Table 2
Basic validity models, traditional and proposed methods (standard errors in parentheses)[a]

	Full sample $n = 18,587$	Eligible subsample $n = 17,346$	
	(A)	(B)	(C[b])
Panel A: both predictors			
HSGPA	0.507	0.571	
	(0.010)	(0.012)	
SAT/1000	0.930	0.928	
	(0.027)	(0.028)	
R^2[c]	0.409	0.454	
R	0.639	0.674	
Panel B: HSGPA only			
HSGPA	0.662	0.726	0.744
	(0.011)	(0.013)	(0.013)
R^2	0.342	0.392	0.396
R	0.585	0.626	0.630
Panel C: SAT only			
SAT/1000	1.485	1.414	1.758
	(0.028)	(0.029)	(0.032)
R^2	0.240	0.228	0.284
R	0.490	0.478	0.533
SAT increment to goodness-of-fit (model A–model B)			
R^2	0.067	0.062	0.058
R	0.055	0.047	0.044

[a]Each column reports three regressions and associated goodness-of-fit statistics. Panel A includes fixed effects for 6 campuses and 18 freshman majors; Panels B and C constrain these effects to be the same as in A.

[b]Restricted models in Panels B and C estimated from Panel A in column B by proposed omitted variables correction.

[c]All fit statistics are corrected for restriction of range, extending models to all 620,013 California SAT-takers. Campus and major effects are excluded from fit calculations. See text for details.

The linear models reported in Fig. 1 are sufficient statistics for calculation of the restricted model (4B). The coefficient of a regression of S on X is $\rho_{SX}\sqrt{\text{var}(S)/\text{var}(X)}$, where ρ_{SX} is the correlation between S and X and all three statistics are calculated from the regression sample. Using the linear projection $\text{E}[S|X] \approx \text{E}[S] + (X - \text{E}[X])\rho_{SX}\sqrt{\text{var}(S)/\text{var}(X)}$, we obtain from (5) the omitted variables formulae:

$$\begin{aligned} \alpha_2 &= \alpha_1 + \left(\text{E}[S] - \text{E}[X]\rho_{SX}\sqrt{\text{var}(S)/\text{var}(X)}\right)\gamma_1, \\ \beta_2 &= \beta_1 + \rho_{SX}\sqrt{\text{var}(S)/\text{var}(X)}\gamma_1, \quad \text{and} \\ \text{var}(y|X) &= \sigma^2 + \gamma_1^2(1-\rho_{SX}^2)\text{var}(S). \end{aligned} \tag{6}$$

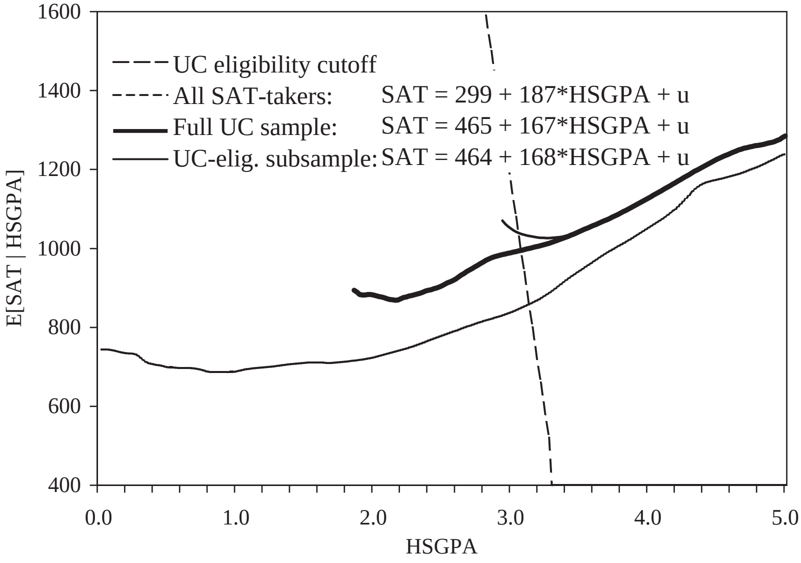

Fig. 1. Conditional expectation of SAT given HSGPA (with linear regression coefficients), three samples. *Note*: Conditional expectation estimated with Epanechnikov kernel (bw = 0.15). Number of observations = 620,013 SAT-takers; 18,587 in full UC sample; 17,346 in eligible subsample. UC sample and subsample HSGPAs incorporate UC bonuses for grades in designated honors courses.

As Fig. 1 indicates, the UC sample inconsistently estimates ρ_{SX}. Eligibility rules that permit a high SAT score to compensate for a low HSGPA lead to a within-sample ρ_{SX} that is lower than its population value. This correlation is 0.52 in the population of SAT-takers, but only 0.43 in the UC sample and 0.38 in the eligible subsample. As a result, the OLS estimates of restricted models in Panels B and C of Table 2 are inconsistent, even when (as in Column B) the sample construction permits consistent estimation of the unrestricted model.

Eq. (6) suggests an estimator for α_2, β_2, and $\text{var}(y|X)$. A consistent estimate of $\rho_{SX}\sqrt{\text{var}(S)/\text{var}(X)}$ can be obtained by regressing S on X in the unselected SAT-taker data. When this is inserted into (6) along with consistent estimates of β_1 and γ_1 from the eligible subsample of the UC data, the resulting estimate of β_2 is consistent. $\text{Var}(y|X)$ can be estimated similarly, using the population data for ρ_{SX} and $\text{var}(S)$. (Note that α_2 is a nuisance parameter for goodness-of-fit statistics, though it could easily be estimated as well.) The same approach estimates γ_3 and $\text{var}(y|S)$, though here the regression in the population data is of X on S.

Column C of Table 2 presents estimates of the restricted models using the omitted variables approach. Coefficients of each are substantially higher than the OLS estimates in Column B.[13] This increases each variable's raw validity and thereby reduces SAT's

[13] The effect is larger for the SAT coefficient in Panel C, again reflecting the high relative weight placed on HSGPAs in eligibility determination.

incremental validity. The usual methods in the full UC sample understate SAT's raw validity by 8% and overstate its incremental validity by 25%, relative to the same statistics estimated by the more defensible methods in Column C.

5. Possible endogeneity of matriculation, campus, and major

The estimates in the rightmost columns of Table 2 correct for eligibility-induced selection-on-unobservables and for inconsistencies in the usual treatment of selection-on-observables, but they do not solve all selection problems. Prediction coefficients still may be biased by selection coming from sources other than eligibility decisions, either from individual campus admissions decisions that select on unobservables or from student matriculation choices between the UC and private alternatives, among UC campuses, and among available majors.[14] In this section, I present alternative specifications meant to assess the bias introduced by these forms of selection. Recall that all eligible students' choice sets include at least one UC campus. I thus treat the student's decision as occurring in two distinct stages, first a choice of UC versus non-UC colleges, then of a campus within the UC system.[15]

Consider first the admitted student's decision about whether to attend the UC. Conditional on HSGPA and SAT, students with high unobserved ability might face better non-UC alternatives than do their peers with similar scores. (Of course, they may also have better within-UC choice sets, as they are likely admitted to more desirable campuses.) One might also imagine that very low ability students have increased costs or reduced benefits of attending UC campuses. Either could induce endogenous sample selection that would bias prediction coefficients toward zero. These stories of endogenous matriculation are most compelling at the extremes of the UC applicant pool, where the UC competes with Stanford and other elite private colleges on the one hand, and with the California State University on the other. Students in the middle are likely to be admitted to a UC campus at approximately their desired selectivity level and to face few comparable alternatives, an effect compounded by the relative rarity of middle-tier private colleges in the western United States.

By this logic, the UC-sample distribution of unobserved characteristics, conditional on observed variables, may be quite truncated at the extremes of the observable distribution, while near the middle of this distribution there is likely to be little selection-on-unobservables. Selection bias should thus be less severe in a trimmed sample that discards observations at the extremes of the UC observable distribution, where the decision to attend the UC may be quite informative about unobserved motivation and ability, in favor of observations near the middle, where the probability of

[14] Decisions about where to submit applications present yet another selection margin. As applications have relatively low cost, however, it seems reasonable to assume that students apply to all colleges at which they have a reasonable probability of admission and at which they would consider matriculating.

[15] I do not consider separately the individual campus admissions decisions, instead allowing these decisions to influence students' choices at each stage through their choice sets. An appendix (available from the author) develops a more complete model of the sample selection process and argues for the estimation strategy taken here.

Table 3
Specification checks (standard errors in parentheses)[a]

	Basic model	Sample trimmed on:			Instrumental variables estimates[b]			
		Elig. index	Pred. FGPA	SAT	Full sample		Trimmed on pred. FGPA	
	(A)	(B)	(C)	(D)	(E)	(F)	(G)	(H)
Prediction coefficients (model A)								
HSGPA	0.57	0.61	0.59	0.58	0.45	0.49	0.44	0.48
	(0.01)	(0.02)	(0.02)	(0.01)	(0.03)	(0.04)	(0.05)	(0.05)
SAT/1000	0.93	0.92	0.95	0.89	0.65	0.71	0.53	0.58
	(0.03)	(0.03)	(0.04)	(0.04)	(0.08)	(0.09)	(0.11)	(0.12)
Campus FEs	y	y	y	y	endog.	endog.	endog.	endog.
Major FEs	y	y	y	y	n	y (4)	n	y (4)
Goodness-of-fit statistics (corrected for restriction of range)[c]								
R^2 A: SAT and HSGPA	0.454	0.472	0.465	0.457	0.259	0.290	0.204	0.229
B: HSGPA only	0.396	0.417	0.407	0.403	0.231	0.258	0.187	0.210
C: SAT only	0.284	0.286	0.289	0.280	0.155	0.173	0.114	0.127
A-B: SAT increment	0.058	0.054	0.058	0.054	0.028	0.031	0.017	0.019
R A: SAT and HSGPA	0.674	0.687	0.682	0.676	0.509	0.538	0.452	0.478
B: HSGPA only	0.630	0.646	0.638	0.635	0.481	0.508	0.432	0.458
C: SAT only	0.533	0.535	0.538	0.529	0.393	0.416	0.337	0.357
A-B: SAT increment	0.044	0.041	0.044	0.041	0.028	0.030	0.019	0.021

[a]Columns A–D estimated by OLS on the UC-eligible subsample; columns E–H by IV. Number of observations = 17,346 in A, E, and F. Columns B–D, G, and H delete observations in top and bottom deciles along listed index, retaining 13,879 observations.

[b]Instruments are indicators for residence in the same county as each of the UC campuses and the continuous distance between the home county and each campus (excluding Irvine).

[c]Fit statistics extend models to 620,013 SAT-takers, using omitted variables estimator and constraining fixed effects in restricted models.

selection into the sample plausibly approaches one. To evaluate whether endogenous matriculation biases the estimates presented in Table 2, models were re-estimated on "trimmed" samples that delete the top and bottom deciles of the eligible subsample.[16] Columns B–D of Table 3 report estimates from samples trimmed along three dimensions: An average of SAT and HSGPA corresponding to the UC eligibility rules, which weight HSGPA heavily; fitted values from the unrestricted model in Table 2, Column B; and the SAT score alone. None of the trimmed samples produces substantially different prediction or validity estimates than does the full sample, suggesting that student matriculation decisions do not much bias validity estimates for eligible students.

[16] I am not aware of previous uses of this sort of specification check, although it has similarities to Altonji et al.'s (2000) approach and to Heckman's (1990) "identification at infinity." The essential insight is that the unobservable distribution is most severely truncated near the selection margin, and that this should be observable in the conditional FGPA distribution. Alternative tests of the same phenomenon might look for changes in the residual variance or in quantile regression slopes near the selection margin.

Having decided to attend the UC, students choose a campus and then a major. Campus assignment is a function both of admissions decisions and of students' own preferences, as the latter must decide to which campuses they will apply and, if accepted to several, at which they will enroll. In the extreme case, the system would offer a continuum of campuses and admissions rules would perfectly stratify students by preparedness; all predictive power from the SAT score would be incorporated into the campus assignment. Within campuses, variation of SAT scores and HSGPAs would be perfectly offset by unobservables and the fixed-effects SAT and HSGPA coefficients would be zero. A similar problem could arise from the use of fixed effects for student major, if students sort into majors based on their ability.

This suggests that both campus and major are potentially endogenously assigned within the UC-eligible subsample. To evaluate this, I estimated the basic fixed-effects model by instrumental variables. It seems likely that students prefer to attend campuses near their homes but that unobserved ability does not vary with geography. Thus, the probability of attending a particular campus might be expected to fall with the distance from that campus and to rise with the distance from other campuses. Fifteen geographic variables were used as instruments for students' campus assignments: eight indicators for residence in the same county as one of the UC campuses and seven measures of the distance between the student's home county and the UC campuses.[17] The full set of major dummies could not be used in the IV model, as certain majors exist only at a single campus and therefore perfectly predict campus assignment.[18] Column E of Table 3 reports IV estimates of a model that excludes major effects, while in Column F majors are collapsed into five broad categories and fixed effects are included for four of them. The instruments are quite powerful predictors of campus assignment, and the first stage coefficients (not reported) generally have the expected signs. F statistics on the exclusion of the instruments in the first stage regressions range from 22 to 201. The IV estimates for the SAT and HSGPA coefficients are notably smaller than in the base model in Column A, with a slightly larger effect on the SAT than the HSGPA. Finally, Columns G and H combine the two strategies, reporting IV models on the trimmed sample. These echo the full-sample IV results.

Taking all the alternative specifications in Table 3 together, the evidence suggests that biases in the basic model in Table 2 are small and, if anything, lead to overstatement of the SAT's role in predictions. The SAT's incremental validity seems at least one fifth smaller (from Table 2) than would be indicated by the usual methods.

[17] The distance to UC Irvine is excluded: Five of the eight campuses are in Southern California, and the eight distance variables are highly collinear. Note that the instruments have no predicted relationship with admissions decisions, only with campus selection conditional on admission. Under the assumption of constant treatment effects, an instrument for student preferences is sufficient. Students not admitted to a particular campus are "never takers," and their probability of attendance is unaffected by the instruments. IV estimates are identified from compliers, students who are admitted to both nearby and faraway campuses and whose choice depends on location (Angrist et al., 1996).

[18] Plausible instruments for major are not apparent. Encouragingly, however, the major effects estimated in Table 2 are similar to more plausibly causal matching results reported by Elliot and Strenta (1988).

6. Decomposing the SAT's predictive power

Thus far, only HSGPA and SAT have been considered as FGPA predictors. This necessarily overstates the predictive accuracy that would be lost were SAT scores unavailable in admissions. The SAT's absence would be partly compensated by re-weighting other predictors like application essays or teacher recommendations.[19] SAT-based and non-SAT-based predictions would be more accurate, and more similar to each other, than is indicated by sparse models.

In this section, I consider the implications of individual- and school-level demographic variables for FGPA prediction, examining whether sparse models overstate the SAT's importance by allowing it to proxy for these demographic characteristics. Specifically, I use the background variables to generate a predicted SAT score for each student in the UC database, and ask whether the predicted score can account for the relationship between SAT scores and FGPAs. If the SAT is a socioeconomically neutral measure of student preparedness, its predictable portion should be no more or less strongly related to FGPA than is the unpredictable portion. On the other hand, to the extent that SAT scores are serving to "launder" the demographic characteristics but do not well measure preparedness conditional on observable student background variables, the fitted SAT score will have a larger coefficient than does the residual.

I consider two categories of student background characteristics. The first consists of individual race (Black, Hispanic, and Asian) and gender indicator variables. The second describes the demographic makeup of the student's high school. Five variables are used: The fraction of students who are Black, Hispanic, and Asian; the fraction of students receiving subsidized lunches; and the average education of students' parents.[20] Because the latter data are available only for public schools, analyses in this section restrict attention to public school students.

The use of school-level predictors is a function of data availability, but also has a substantive justification: The SAT has long been advocated as a necessary check on potentially heterogeneous high school grading policies (Caperton, 2001), and the College Board argues that admissions rules which consider only variables under the high school's control induce high school grade inflation. To the extent that the SAT's role is to limit such a tendency, it might be expected to have more predictive power across high schools than within. The implications of this, however, depend on whether observable characteristics of high schools can serve the same role. If "grade inflation" is highly correlated with socioeconomic variables, HSGPAs can be appropriately discounted without access to SAT scores.

Columns A–C of Table 4 present several OLS regressions with SAT scores as the dependent variable. Individual and school background characteristics are strong

[19] Willingham (1985) investigates the supplementary application variables that are used in admissions but are unavailable in the UC data. He finds them to be significant predictors of college success when SAT is controlled, but does not report their effect on the SAT's incremental contribution.

[20] The results are insensitive to the exclusion of average parental education, the least widely available background variable, and also to the inclusion of the school's exam-based API score as a measure of school effectiveness. It would be desirable to test the importance of individual parental education, but this is not observed in the UC data.

Table 4
Individual and school characteristics as determinants of SAT scores and GPAs (standard errors in parentheses)[a]

Sample	SAT-takers				UC sample (UC elig. only)		
Dependent variable	SAT/1000			HSGPA	SAT/1000	HSGPA	FGPA[b]
	(A)	(B)	(C)	(D)	(E)	(F)	(G)
Intercept	0.979		0.049	3.011	0.488	4.231	2.532
	(0.001)		(0.008)	(0.026)	(0.039)	(0.095)	(0.156)
Black	−0.220		−0.151	−0.403	−0.105	−0.244	−0.291
	(0.001)		(0.001)	(0.004)	(0.008)	(0.020)	(0.033)
Hispanic	−0.175		−0.104	−0.186	−0.105	−0.152	−0.248
	(0.001)		(0.001)	(0.003)	(0.005)	(0.011)	(0.019)
Asian	−0.026		−0.010	0.134	−0.012	0.091	−0.067
	(0.001)		(0.001)	(0.002)	(0.003)	(0.008)	(0.013)
Female	−0.044		−0.040	0.120	−0.068	0.036	0.061
	(0.001)		(0.001)	(0.002)	(0.003)	(0.006)	(0.010)
Skl: Frac. Black		−0.253	−0.119	−0.159	−0.076	−0.039	−0.389
		(0.003)	(0.003)	(0.010)	(0.015)	(0.037)	(0.061)
Skl: Frac. Hispanic		−0.023	0.057	−0.084	0.061	−0.021	−0.099
		(0.003)	(0.003)	(0.008)	(0.012)	(0.031)	(0.050)
Skl: Frac. Asian		0.041	0.037	−0.272	0.061	−0.184	0.195
		(0.002)	(0.002)	(0.006)	(0.008)	(0.021)	(0.034)
Skl: Frac. free lunch		0.001	0.006	0.037	−0.022	−0.048	−0.017
		(0.003)	(0.003)	(0.008)	(0.012)	(0.030)	(0.049)
Skl: Avg. parental ed.		0.062	0.063	0.020	0.045	−0.026	0.033
		(0.001)	(0.001)	(0.002)	(0.002)	(0.006)	(0.010)
R^2	0.134[c]	0.177	0.225	0.069	0.230	0.050	0.053

[a]Students from public high schools with non-missing data only. Number of observations = 473,758 in columns A–D; 14,102 in E–G.
[b]FGPA is adjusted to remove estimated campus and major effects from Table 1, column D.
[c]Reported R^2 is the traditional measure, estimated within-sample and unadjusted for restriction of range.

predictors of SATs, and together account for over one fifth of their variance in the California public school SAT-taking population. Column D presents an analogous model for HSGPA on the same sample. Although coefficients on the background variables are all statistically significant in this model, they account for a much smaller share of the HSGPA variance than of SATs. Finally, Columns E–G present models estimated on the sample of eligible public-school UC students. These again indicate that SAT scores are highly correlated with student background, much more so than are either HSGPAs or FGPAs. Other notable differences between the models are that female students have higher HSGPAs and FGPAs but lower SATs than males, and that Asians have higher HSGPAs but lower SATs and FGPAs than whites.

The decompositions in Table 4 suggest that the SAT's role in prediction models may be quite sensitive to the inclusion of background variables, particularly school characteristics, as predictors. School and individual demographic characteristics explain fully 23% of the SAT variance in the UC sample, but account for only 5% of variance of each of the grade-based variables. Models which replace the school characteristics with high school fixed effects, not reported in Table 4, indicate that observable demographics account for a substantial share of the across-school variation in SAT scores.

Table 5 considers whether the SAT's heavy loading onto student background characteristics accounts for its predictive power for FGPA. Column A repeats the sparse FGPA model from Table 2 using only eligible students from public schools. Columns B–D include as an additional FGPA predictor the fitted SAT score from Columns A–C, respectively, of Table 4.

These models indicate that characteristics of students' schools, though not individual race and gender, account for a large share of the SAT's predictive power. Consider Column D, which effectively decomposes SAT scores into a portion that reflects student and school characteristics and an individual innovation. Two students who differ in background characteristics producing a 100-point gap in fitted SAT scores (a student from an all-white school and an otherwise identical student from a school that is nearly all black, for example), earn FGPAs that differ, on average, by 0.13. If we instead compare two students with identical observable characteristics but SAT *residuals* that differ by 100 points—i.e. students with the same predicted SATs but different actual SATs—the expected gap in FGPAs is only 0.07 points. The latter, the SAT coefficient when its easily observed correlates are controlled, characterizes the independent information provided by SAT scores. The sparse models used in the literature conflate the two SAT portions and predict a 0.09 point gap regardless of the source of SAT differences.

Columns E–G carry out a similar exercise, this time allowing student background characteristics to predict FGPAs directly. F tests reject the restrictions imposed in the earlier models at any reasonable confidence level, but the new models do not change the substantive interpretation: SAT scores are less informative, net of the information they provide about student background, about FGPAs than is implied by sparse models. Coefficients on the background variables are generally what one might expect, with racial minorities and students from schools with high concentrations of Blacks, Hispanics, or low-education parents earning lower FGPAs than white, upper-SES students even when SAT scores and HSGPAs are controlled. As other authors have found, women earn higher FGPAs than expected given their HSGPAs and SATs (see, e.g., Leonard and Jiang, 1999).[21] One coefficient is somewhat surprising: Asian students earn lower FGPAs than do otherwise similar Whites, although students from schools with many Asian students do quite well (see also Young, 2001).

There is a sense in which the inclusion of variables not used for admissions in Table 5 can lead to understatement of the SAT's role. The two SAT subscores have reliabilities of about 0.9 (College Board, 2001). Saturation of prediction models concentrates the unreliability, attenuating the SAT coefficient and inflating coefficients on

[21] This may account for the relatively small coefficient on the fitted SAT in Column B: The racial component of the fitted SAT appears strongly related to FGPA while the gender component seems negatively correlated.

Table 5
Individual and school characteristics in FGPA prediction (standard errors in parentheses)[a]

	(A)	(B)	(C)	(D)	(E)	(F)	(G)
HSGPA	0.583	0.581	0.619	0.602	0.554	0.621	0.604
	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)
	[0.556][b]	[0.554]	[0.595]	[0.578]	[0.526]	[0.597]	[0.578]
SAT/1000	0.906	0.875	0.694	0.705	0.886	0.694	0.725
	(0.031)	(0.034)	(0.034)	(0.036)	(0.034)	(0.034)	(0.036)
	[1.050]	[1.032]	[0.832]	[0.852]	[1.047]	[0.833]	[0.880]
Demographic measures							
Predicted SAT/1000[c]		0.187	0.818	0.635			
		(0.078)	(0.055)	(0.056)			
		[0.066]	[0.698]	[0.511]			
Black					−0.136		−0.073
					(0.030)		(0.030)
Hispanic					−0.125		−0.082
					(0.017)		(0.017)
Asian					−0.094		−0.122
					(0.010)		(0.011)
Female					0.110		0.100
					(0.010)		(0.010)
Skl: Frac. Black						−0.326	−0.296
						(0.053)	(0.054)
Skl: Frac. Hispanic						−0.195	−0.146
						(0.045)	(0.045)
Skl: Frac. Asian						0.145	0.265
						(0.028)	(0.030)
Skl: Frac. free lunch						0.032	0.026
						(0.044)	(0.043)
Skl: Avg. Parental Ed.						0.021	0.016
						(0.009)	(0.009)
R^2 (range corrected)[d]	0.459	0.457	0.471	0.462	0.461	0.475	0.481
Without SAT	0.404	0.410	0.445	0.435	0.414	0.449	0.454
SAT Increment	0.056	0.047	0.027	0.027	0.048	0.026	0.027
R (range corrected)	0.678	0.676	0.687	0.680	0.679	0.689	0.694
Without SAT	0.635	0.641	0.667	0.660	0.643	0.670	0.674
SAT Increment	0.042	0.035	0.020	0.020	0.036	0.019	0.020

[a]Sample in all columns consists of UC-eligible, public school graduates with non-missing data. Number of observations = 14,102. All models include fixed effects for 6 campuses and 18 freshman majors.

[b]Errors-in-variables corrected coefficients, assuming SAT reliability = 0.9, in square brackets. Corrected demographic variable coefficients are not shown.

[c]Predicted SAT is fitted value from regressions reported in Table 4, columns A–C, respectively.

[d]Goodness-of-fit statistics extend results to 473,758 public school SAT-takers. “Without SAT” statistics are based on unreported restricted models that exclude the SAT score and are calculated by omitted variables estimator described in text.

variables correlated with SAT scores even more than in sparse models. One would not want to correct ordinary validity estimates for this—admissions offices do not have access to a perfectly reliable SAT score, only to the noisy one—but the saturated models risk overstating the coefficient on the fitted SAT relative to that on the actual SAT.

Table 5 reports, in square brackets, selected coefficients derived from a multivariate errors-in-variables correction (Greene, 2000, p. 378), assuming a SAT reliability ratio of 0.9.[22] The adjusted coefficients indicate that the decline in the SAT coefficient when individual race and gender are controlled is primarily a statistical artifact, but do not affect the interpretation of the results for school characteristics.

The final rows of Table 5 present range-corrected goodness-of-fit statistics, based on uncorrected coefficients, for both the full models and omitted-variables specifications that exclude the actual SAT score but retain the predicted score or demographic variables. The latter measures might be used in predictions even if SAT scores were not, and their predictive power is thus not attributable to the SAT score. The inclusion of school demographic controls—either separately or through the predicted SAT score—lowers the SAT's estimated incremental validity by about 50%.

7. Discussion

This study has addressed two methodological concerns generally ignored in the SAT validity literature. First, it has embedded the sample selection problem in an explicit model, proposing a new estimator that is consistent under certain assumptions. These assumptions are not general—the current analysis benefits from the UC's unusual reliance on easily observed characteristics for eligibility decisions—but are more reasonable than the internally inconsistent assumptions needed to support usual practice. Researchers working with data from other colleges will typically not be able to rely on the selection-on-observables assumption that permits the current analysis, but might consider using the "trimming" approach from Section 5 along with the omitted variables estimator to assess the magnitude of selection biases. I estimate that the usual methods overstate the SAT's incremental validity for University of California FGPAs by about one quarter relative to the selection-adjusted estimate. Additional specifications do not indicate substantial downward bias in the latter from non-eligibility-based forms of sample selection.

The selection results are interesting, but not particularly informative about admissions policy. After all, an incremental validity of 0.044 may well be enough to justify use of the SAT. The second portion of the analysis, focusing on the role of demographic variables in FGPA prediction, is of more direct substantive interest.

The results in Tables 4 and 5 suggest that in sparse models the SAT serves in part to proxy for student background characteristics. These variables account for a substantial share of the variance in SAT scores. They are also strong predictors of FGPA in their own right—together with HSGPA, school and individual demographic variables explain 45% of the variance in FGPAs, about as much as do SAT and HSGPA together in models excluding background variables. Moreover, fitted SATs predicted from student

[22] This is probably a lower bound for the SAT composite's reliability, as true scores are likely to be more highly correlated across the two subtests than is noise.

and school demographic variables are more strongly related to FGPAs than are actual SATs.[23]

Table 5 indicates that admissions offices could admit better-prepared entering classes by giving explicit admissions preferences to high-SES students and to students from high-SES high schools.[24] SAT scores would receive some weight in "best predictor" admissions rules, but considerably less than is indicated by sparse models.

Few would advocate this sort of admissions rule, which might be called "affirmative action for high SES children," and even fewer would consider it meritocratic. A decision not to consider student background characteristics explicitly in prediction models used for admissions, however, does not justify excluding them from SAT validation models. If background characteristics are not accounted for, the researcher will assign predictive power to any variable that correlates with the excluded variables, whether or not it conveys independent information about preparedness. The results here indicate that the exclusion of student background characteristics from prediction models inflates the SAT's apparent validity, as the SAT score appears to be a more effective measure of the demographic characteristics that predict UC FGPAs than it is of variations in preparedness conditional on student background. A policy-maker who preferred not to use demographic variation to identify students likely to succeed might want to build an admissions rule around the SAT and HSGPA coefficients in Table 5, while ignoring the coefficients on demographic control variables.[25]

Regardless of one's view of the appropriate role of student background characteristics in admissions, the results here suggest that the SAT should be assigned less importance than is implied by the sparse, selection-biased models in the validity literature. If one wishes to exploit the predictive power of student background, the background variables themselves can provide much of the information contained in the SAT score; if one does not wish to use background in prediction, the SAT's contribution is smaller than the validity literature would suggest. Comparing incremental validities in Tables 2 (Column A) and 5 (Column D), a conservative estimate is that traditional methods and sparse models overstate the SAT's importance to predictive accuracy by 150%.

Acknowledgements

I am grateful to Chris Avery; Jared Bernstein; David Card; Ken Chay; David Lee; Darren Lubotsky; Justin McCrary; Jack Porter; Eugene Smolensky; seminar participants at Berkeley, Princeton, NBER, and the UCSB *Rethinking the SAT* conference; and

[23] One interpretation is that the SES gradient in SAT scores is too low: An increase in this gradient could permit the predictable and unpredictable parts of the SAT to have the same coefficient in FGPA models. The high explanatory power of models for SAT in Table 4, however, suggests an arguably more reasonable interpretation: the SAT score captures background characteristics more than it independently measures student preparedness.

[24] Note that affirmative action typically assigns preferences in the opposite direction from that indicated by Table 5.

[25] A more extreme reaction would be to admit using only the residuals from models like those in Table 4, rather than the entire HSGPA and SAT (Studley, 2001). Percent plans, which base admissions on within-school rank in class, do essentially this.

three anonymous referees for helpful suggestions. I also thank Saul Geiser and Roger Studley at the University of California Office of the President and David Card and Alan Krueger for access to data. This material is based upon work supported under a National Science Foundation Graduate Research Fellowship and a grant from the University of California's Center for Studies in Higher Education.

References

Altonji, J.G., Elder, T.E., Taber, C.R., 2000. Selection on observed and unobserved variables: assessing the effectiveness of Catholic schools. NBER Working Paper 7831.

Angrist, J.D., Imbens, G.W., Rubin, D.B., 1996. Identification of causal effects using instrumental variables. Journal of the American Statistical Association 91, 444–455.

Atkinson, R.C., 2001. Standardized tests and access to American universities. Atwell Lecture at 83rd Annual meeting of the American Council of Education, Washington, DC.

Barro, R.J., 2001. Economic viewpoint: why colleges shouldn't dump the SAT. Business Week, April 9, 20.

Bowen, W.G., Bok, D., 1998. The Shape of the River: Long-term Consequences of Considering Race in College and University Admissions. Princeton University Press, Princeton, NJ.

Breland, H.M., 1979. Population validity and college entrance measures. College Entrance Examination Board, New York.

Bridgeman, B., McCamley-Jenkins, L., Ervin, N., 2000. Predictions of freshman grade-point average from the revised and recentered SAT I: Reasoning test, Research Report 2000–1. College Entrance Examination Board, New York.

Camara, W.J., 2001. There is no mystery when it comes to the SAT I. College Board News.

Camara, W.J., Echternacht, G., 2000. The SAT I and high school grades: utility in predicting success in college. Research Report RN-10, College Entrance Examination Board, New York.

Caperton, G., 2001. Toward a holistic definition of merit: a response to the UC proposal. College Board News.

College Board, 2001. College Bound Seniors 2001. College Entrance Examination Board, New York.

Crouse, J., Trusheim, D., 1988. The Case Against the SAT. University of Chicago Press, Chicago.

Elliot, R., Strenta, A.C., 1988. Effects of improving the reliability of the GPA on prediction generally and on comparative predictions for gender and race particularly. Journal of Educational Measurement 25, 333–347.

Geiser, S., Studley, R., 2001. UC and the SAT: predictive validity and differential impact of the SAT I and SAT II at the University of California. Manuscript, UC Office of the President.

Goldman, R.D., Widawski, M.H., 1976. A within-subjects technique for comparing college grading standards: implications in the validity of the evaluation of college achievement. Educational and Psychological Measurement 36, 381–390.

Greene, W.H., 2000. Econometric Analysis, 4th Edition. Prentice-Hall, Englewood Cliffs, NJ.

Heckman, J., 1990. Varieties of selection bias. American Economic Review 80, 313–318.

Hernandez, M.A., 1997. A Is For Admission: The Insider's Guide to Getting Into The Ivy League And Other Top Colleges. Warner Books, New York.

Leonard, D.K., Jiang, J., 1999. Gender bias and the college predictions of the SATs: a cry of despair. Research in Higher Education 40, 375–407.

McWhorter, J.H., 2001. Eliminating the SAT could derail the progress of minority students. Chronicle of Higher Education, March 9, B11–B12.

Rothstein, J., 2002. College performance predictions and the SAT. Working Paper #45, UC Berkeley Center for Labor Economics.

Stricker, L.J., 1991. Current validity of 1975 and 1985 SATs: implications for validity trends since the mid-1970s. Journal of Educational Measurement 28, 93–98.

Studley, R., 2001. Inequality and college admissions. Unpublished Manuscript.

United States Commission on Civil Rights, 2000. Toward an understanding of percentage plans in higher education: are they effective substitutes for affirmative action? Washington, DC.

University of California Office of the President, 1993. Introducing the University of California: information for prospective students.

Willingham, W.W., 1985. Success in college: the role of personal qualities and academic ability. College Entrance Examination Board, New York.

Willingham, W.W., Breland, H.M., 1982. Personal qualities and college admissions. College Entrance Examination Board, New York.

Willingham, W.W., Lewis, C., Morgan, R., Ramist, L., 1990. Predicting college grades: an analysis of institutional trends over two decades. Educational Testing Service, Princeton, NJ.

Wilson, K.M., 1983. A review of research on the prediction of academic performance after the freshman year. College Entrance Examination Board, New York.

Young, J.W., 1993. Grade adjustment methods. Review of Educational Research 63, 151–165.

Young, J.W., 2001. Differential validity, differential prediction, and college admissions testing. Presentation at Rethinking the SAT in University Admissions Conference, University of California, Santa Barbara.