



Substituting SAT[®] II: Subject Tests for SAT I: Reasoning Test: Impact on Admitted Class Composition and Quality

Brent Bridgeman, Nancy Burton,
and Frederick Cline

Substituting SAT[®] II: Subject Tests for SAT I: Reasoning Test: Impact on Admitted Class Composition and Quality

Brent Bridgeman, Nancy Burton,
and Frederick Cline

Brent Bridgeman is principal research scientist at Educational Testing Service.

Nancy Burton is senior research scientist at ETS.

Frederick Cline is associate research data analyst at ETS.

Researchers are encouraged to freely express their professional judgment. Therefore, points of view or opinions stated in College Board Reports do not necessarily represent official College Board position or policy.

The College Board is a national nonprofit membership association dedicated to preparing, inspiring, and connecting students to college and opportunity. Founded in 1900, the association is composed of more than 3,900 schools, colleges, universities, and other educational organizations. Each year, the College Board serves over three million students and their parents, 22,000 high schools, and 3,500 colleges through major programs and services in college admission, guidance, assessment, financial aid, enrollment, and teaching and learning. Among its best-known programs are the SAT[®], the PSAT/NMSQT[™], the Advanced Placement Program[®] (AP[®]), and Pacesetter[®]. The College Board is committed to the principles of equity and excellence, and that commitment is embodied in all of its programs, services, activities, and concerns.

Additional copies of this report (item #991380) may be obtained from the College Board, 45 Columbus Avenue, New York, NY 10023-6992, 212 713-8165. The price is \$15. Please include \$4 for postage and handling. Purchase orders above \$25 are accepted.

Copyright © 2001 by College Entrance Examination Board. All rights reserved. College Board, Advanced Placement Program, AP, Pacesetter, SAT, and the acorn logo are registered trademarks of the College Entrance Examination Board. PSAT/NMSQT is a joint trademark owned by the College Entrance Examination Board and the National Merit Scholarship Corporation. Other products and services may be trademarks of their respective owners. Visit College Board on the Web: www.collegeboard.com.

Printed in the United States of America

Contents

| | |
|--|---|
| <i>Abstract</i> | 1 |
| <i>Introduction</i> | 1 |
| <i>Method</i> | 1 |
| <i>Sample</i> | 1 |
| <i>Analyses</i> | 2 |
| <i>Results and Discussion</i> | 3 |
| <i>Selecting the Top One-third</i> | 6 |
| <i>Conclusion</i> | 9 |
| <i>References</i> | 9 |
| <i>Tables</i> | |
| 1. Success (GPA 2.5 or greater) Rates for Students Selected or Rejected in Top $\frac{2}{3}$ by HVM and/or H(SA) | 3 |
| 2. Success (GPA 2.5 or greater) Rates for Students Selected or Rejected in Top $\frac{1}{3}$ by HVM and/or H(SA) | 8 |

Figures

| | |
|--|---|
| 1. Percent of females in top $\frac{2}{3}$ selected under six alternative models..... | 4 |
| 2. Percent of each group selected by HVM, H(SA), and H(SA-NL) for upper $\frac{2}{3}$ | 4 |
| 3. Percent of each ethnic/ESL subgroup selected by HVM, H(SA), and H(SA-NL) for upper $\frac{2}{3}$ | 5 |
| 4. For students who took a Spanish Subject Test, percent of each group selected by HVM, H(SA), and H(SA-NL) for upper $\frac{2}{3}$ | 6 |
| 5. For students who took a Spanish Subject Test, percent of selected students in each group who are successful (GPA 2.5 or higher) | 7 |
| 6. For students who took a Spanish Subject Test, number of selected students in each group who are successful (GPA 2.5 or higher) | 7 |
| 7. Percent of each group selected by HVM, H(SA), and H(SA-NL) for upper $\frac{1}{3}$ | 8 |

Abstract

Using data from a sample of 10 colleges at which most students had taken both SAT® I: Reasoning Test and SAT II: Subject Tests, we simulated the effects of making selection decisions using SAT II scores in place of SAT I scores. Specifically, we treated the students in each college as forming the applicant pool for a more select college, and then selected the top two-thirds (and top one-third) of the students using high school grade point average (HSGPA) combined with either SAT I scores or the average of SAT II scores. Success rates, in terms of freshman grade point averages, were virtually identical for students selected by the different models. The percent of African American, Asian American, and White students selected varied only slightly across models. Appreciably more Mexican American and Other Latino students were selected with the model that used SAT II scores in place of SAT I scores because these students submitted Subject Test scores for the Spanish Test on which they had high scores.

Key words: SAT II validity, achievement versus aptitude, selection models

Introduction

The SAT I: Reasoning Test measures “verbal and mathematical reasoning abilities, which develop over time” (College Board, 1999a, p.3). The SAT II: Subject Tests “measure your knowledge and skills in particular subjects and your ability to apply that knowledge” (College Board, 1999b, p.3). In terms of their overall ability to predict freshman grades, the SAT I and SAT II tests may be nearly identical. Using data from 22 highly selective colleges that used the SAT and Achievement tests (predecessors to the SAT I and SAT II), Crouse and Trusheim (1988) found essentially no difference in the ability of these two types of tests to predict freshman grade point average (FGPA). This conclusion of no difference held whether the tests were used by themselves or combined with high school grades. They suggest that factors other than the ability to predict FGPA may then enter into decisions of which type of test should be used.

One argument for using achievement tests rather than general developed ability tests is the simple assertion that general tests are biased and unfair (McClelland, 1973; see Barrett and Depinet, 1991, for a critical assessment of McClelland’s assertions and McClelland, 1994, for his response to Barrett and

Depinet). Because of their closer link to school subjects, though not to a particular well-specified curriculum, SAT II tests may be seen as inherently less vulnerable to complaints of test bias. Poor performance on SAT II: Chemistry, for example, is more likely to be attributed to the quality of the chemistry instruction in the school or the student’s work in chemistry class rather than test bias. A different argument suggests that even tests that are not a direct measure of the curriculum will influence instruction and thus should contain content that is worth being taught (Linn, 1994; Messick, 1989; Resnick and Resnick, 1992; Shepard, 1992, 1997).

Colleges in the University of California system now weight SAT II more heavily than SAT I for making admission decisions. For example, the University of California: San Diego uses the following equation to rank students: $HSGPA \times 1000 + [(SAT\ I\ Verbal + SAT\ I\ Math + SAT\ II\ Writing + SAT\ II\ Math + SAT\ II\ third\ test) \times .8]$ (UCSD, 2000), and some have proposed simply substituting SAT II for SAT I (e.g., Crouse and Trusheim, 1988). Such a substitution could impact not only the quality of the class selected, but also its gender and ethnic composition.

The current research focuses on the consequences of substituting SAT II tests for the SAT I in the selection of a freshman class. Because the database to be used includes freshman grades, we can model not only the composition of the selected class, but also its academic success, at least to the extent that success can be defined by grades. This modeling is necessarily limited to the data available, namely test scores and grades. We do not intend to suggest that these indicators are or should be the only factors considered in making admission decisions. Nevertheless, as long as test scores remain one of the important considerations in selective admission, any differences produced by the tests in the nature and composition of the students admitted are relevant.

Method

Sample

Colleges in the sample were selected from a database of 23 colleges that was assembled for an SAT I validity study that compared the predictive validity of the old SAT to the new SAT I (Bridgeman, McCamley-Jenkins, & Ervin, 2000). This database contains SAT I and SAT II scores and responses on the Student Descriptive Questionnaire (SDQ), including student reported high

school grade point average, ethnic identification, best language (English, English and another, or another), parental education, family income, and intended college major. In addition, the database contains the freshman grade point average (FGPA). Students in the database were freshmen in 1995, so scores were available for relatively recent versions of the SAT II tests including Writing and Math IIC (advanced math that requires calculator use).

From this database of admitted and enrolled students, we selected only colleges in which at least 80 percent of the freshman class had taken SAT II: Writing plus at least one other SAT II: Subject Test. Thus, at the campuses studied, students who took SAT II tests were the rule and not the exception. The 10 colleges included in the final sample were: Barnard, Bowdoin, Colby, Harvard, Northwestern, four campuses of the University of California (Davis, Irvine, Los Angeles, and San Diego), and Vanderbilt. Mean SAT I: verbal scores in these colleges ranged from 524 to 731 and mean SAT I: math scores ranged from 574 to 726; in all but two of the colleges, mean verbal and math scores were above 600. The range of scores within each institution was more restricted than in the national sample, but there was still substantial within-college variation with standard deviations ranging from 58 to 92 (compared to national standard deviations of about 110 for both verbal and math scores).

Responses to the Student Descriptive Questionnaire that students fill out when they register for the SAT were used for ethnic group identification. Based on these self reports, the sample contained 500 African American, 4,725 Asian American, 923 Mexican American, 542 Other Latino, and 6,086 White students. There were 6,264 male and 7,610 female students.

Analyses

Freshmen at each of the 10 colleges who had scores on SAT II: Writing Test and at least one other SAT II test were treated as if they formed an applicant pool for an even more selective institution. At each college, two-thirds of the “applicant pool” was “selected” based on various score composites. A second set of analyses “selected” the top one-third. Because any realistic selection scenario would include the high school grade point average (H), we decided to include H in each composite even though this would have the effect of muting the differences between selections made by alternative models. The self-reported H scores in this select sample had a narrow range with no student reporting an average lower than C, and 92 percent of the students reported grade averages in one of the four

highest categories (B+ through A+). These grades were placed on an SAT-like scale by setting a C to 400 and proceeding in 50-point increments to 750 for an A+, producing a scale with a mean of 669 and a standard deviation of 61. Within-college standard deviations for H ranged from 40 to 63. In each model, the composite score was formed by equally weighting each test score and giving H nominally equal weight with the combined test scores (e.g., if there were two SAT I scores [verbal and mathematical] and one SAT II score, these scores would be summed and H would be multiplied by 3 and added to the total). The technique of using data on enrolled students to model the results of employing different admission strategies has been used for many years (see, for example, Kane, 1998; Wightman, 1997; Willingham and Breland, 1982; and Wing and Wallach, 1971).

The following composites were used:

(Note: H=high school GPA; V=SAT I-Verbal; M=SAT I-Math; W=SAT II: Writing)

H+V+M (HVM)

H+Subject Test Average (H[SA])

H+Subject Test Average excluding language tests (H[SA-NL])

H+V+M+W (HVMW)

H+V+M+best Subject Test (HVMB)

H+V+M+best non-language Subject Test (HVM[B-NL])

The averages that excluded language tests were included in the analyses because of the possibly unique role that language tests could play. Most Subject Tests are measures of school learning, but when language tests are taken by native speakers of those languages, they are measures primarily of out-of-school learning.

Students selected by one of the new composites with SAT II scores were compared to students selected by the traditional HVM index. We defined successful students as those who attained a freshman grade point average of at least 2.5. (We also investigated a freshman GPA of 2.0 or better as the criterion, but the overall success rate was 87 percent, allowing for little variation among the different selection methods.) The percent of successful students selected was compared for four groups: (1) students selected by both the new composite and traditional index, (2) students rejected by both methods, (3) students selected by the new but rejected by the traditional, and (4) students selected by the traditional but rejected by the new.

Results and Discussion

Table 1 compares the percent of students who were successful when selected by HVM to the percent who were successful when selected by H plus the average of the Subject Tests (H[SA]). Because of the high correlation between SAT I and the average of the SAT II tests ($r = .84$ for the total sample), and because H was used in both selection methods, the selection decision was the same under both models for 86 percent of the students. The comparison of Group 1 (students selected by both procedures) and Group 2 (students rejected by both procedures) suggests that valid selections can be made even though the initial selection pools were already quite restricted because they consisted of only students who had already been admitted to and enrolled in selective colleges. The comparison of Groups 3 and 4 (the 14 percent of the students who were selected by one method but rejected by the other) indicated that the percentage of successful students in these two groups was nearly identical.

A complementary analysis that focused on grade averages rather than percent of successful students reached the same conclusion. In this analysis, standardized differences in FGPA (difference divided by the weighted within group standard deviation) were computed within each college comparing the FGPA of students in Group 1 with students in Group 2 and comparing students in Group 3 with students in Group 4. These standardized differences were weighted by the number of students in the relevant groups in each college and averaged across colleges. The standardized difference between groups 1 and 2 was 0.82 (standard error = 0.06) indicating that grades of students selected by both methods were substantially above the grades of students rejected by both methods. The difference between groups 3 and 4 was only 0.03 (standard error = 0.05). Thus, with freshman grades as the criterion,

there is no reason to favor either SAT I or SAT II in making selection decisions. The same was true for the comparison that excluded language tests from the Subject Test average.

As expected, there was even more overlap in the models that added Subject Tests to V and M rather than replacing V and M. In the model that added Writing, 93 percent of the selection decisions were the same as with HVM alone. Similarly, for HVMB (adding the best Subject Test score to the SAT I scores and high school grades), 93 percent of the decisions were identical, and for HVM(B-NL), 94 percent were identical. There were no significant differences in any of the comparisons between the FGPA's of the groups admitted by one model and rejected by the other.

Although the overall success of students selected using SAT I is comparable to the success of students selected using SAT II, there might still be differences in the ethnic or gender composition of groups selected by the different criteria. Figure 1 shows the percent of the pool of female students that was selected by being in the top two-thirds for each selection index. Each percentage was slightly below the 66.7 percent that would be expected if there were no gender differences on any of the selection instruments. Although there was relatively little variation among the various indices, including SAT II: Writing Test along with SAT I scores increased the percentage of women selected by a small but statistically significant 2.5 percentage points (standard error of each percentage is about 0.7).

As indicated in Figure 2, more substantial differences were evident in the ethnic group comparison of HVM selections with selections that combined H with the average of the Subject Tests (H[SA]) and selections that combined H with the average of the non-language Subject Tests (H[SA-NL]). In particular, the proportion of Mexican American and Other Latino students selected would increase if H(SA) were used in place of HVM. Because we were keeping the size of the admitted

TABLE 1

Success (GPA 2.5 or greater) Rates for Students Selected or Rejected in Top $\frac{2}{3}$ by HVM and/or H(SA)

| Group | Group 1 (in both) | | Group 2 (out both) | | Group 3 (in H[SA] only) | | Group 4 (in HVM only) | |
|--------------|-------------------|------------|--------------------|------------|-------------------------|------------|-----------------------|------------|
| | n | % GPA 2.5+ | n | % GPA 2.5+ | n | % GPA 2.5+ | n | % GPA 2.5+ |
| Men | 3,916 | 84 | 1,508 | 61 | 380 | 71 | 460 | 72 |
| Women | 4,318 | 90 | 2,175 | 71 | 609 | 78 | 508 | 81 |
| African Am. | 119 | 88 | 326 | 65 | 27 | 85 | 28 | 75 |
| Asian Am. | 2,895 | 84 | 1,147 | 63 | 329 | 71 | 354 | 69 |
| Mexican Am. | 280 | 80 | 461 | 56 | 140 | 59 | 42 | 60 |
| Other Latino | 226 | 86 | 201 | 58 | 93 | 72 | 29 | 72 |
| White | 4,054 | 90 | 125 | 75 | 327 | 85 | 448 | 85 |
| Total | 8,234 | 87 | 3,683 | 67 | 989 | 75 | 968 | 77 |

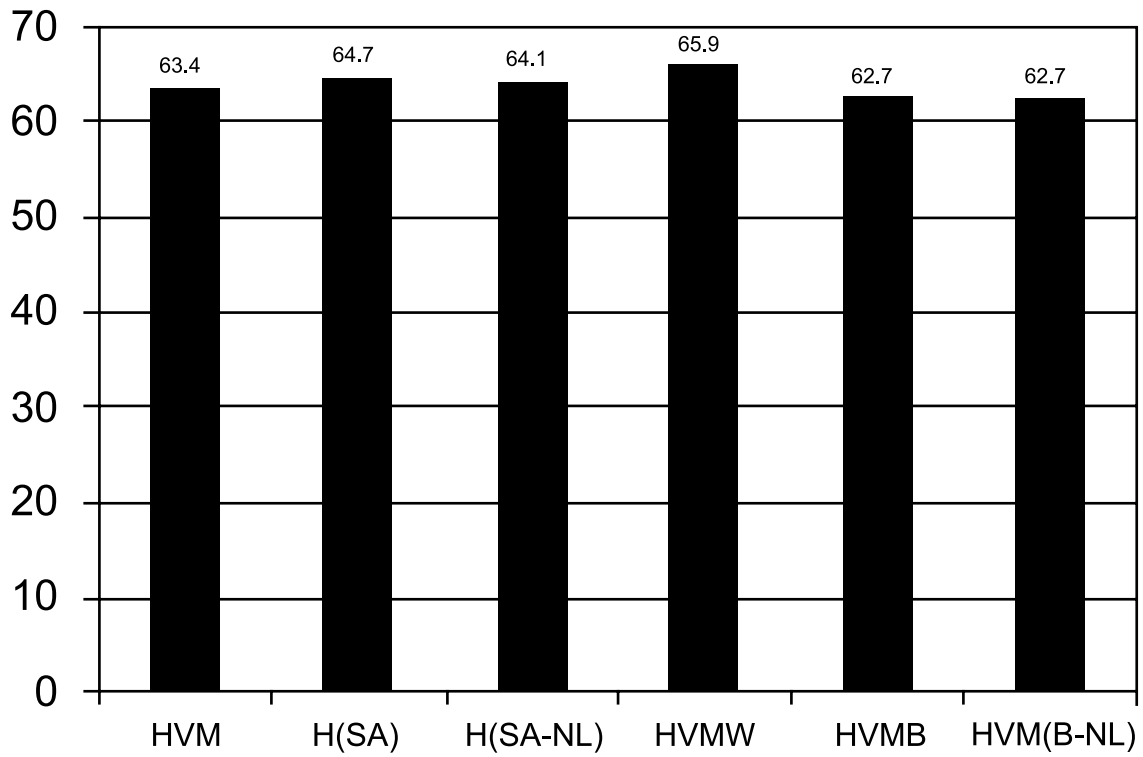


Figure 1. Percent of females in top $\frac{2}{3}$ selected under six alternative models.

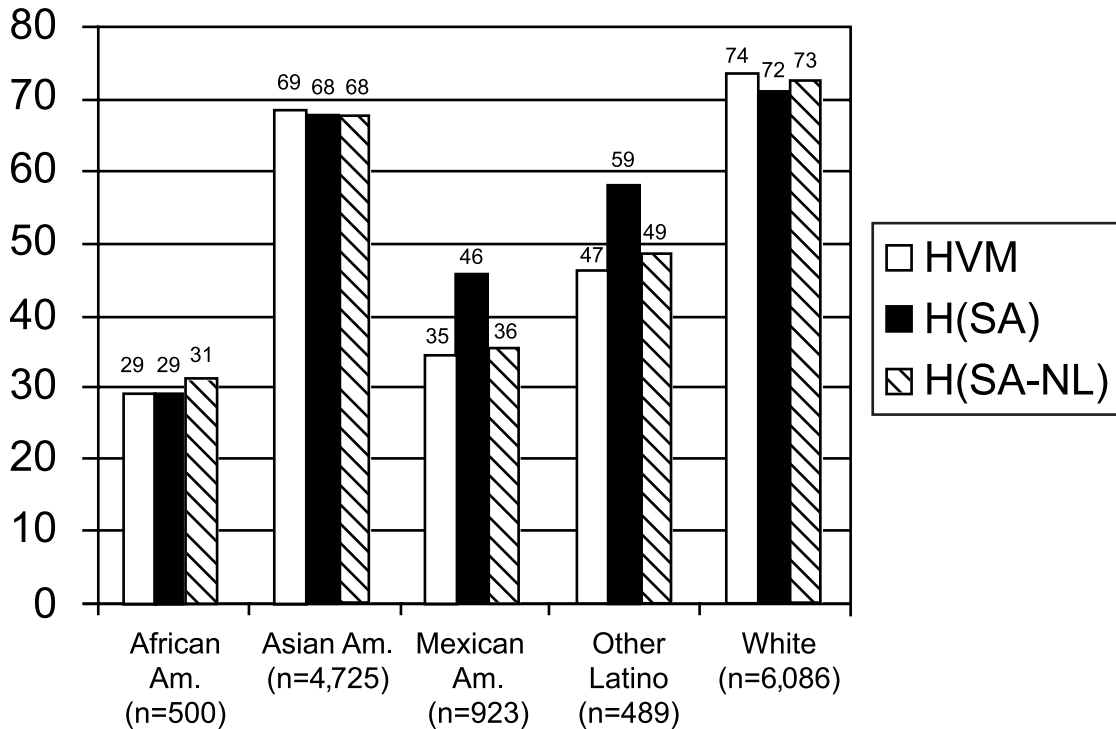


Figure 2. Percent of each group selected by HVM, H(SA), and H(SA-NL) for upper $\frac{2}{3}$.

class fixed, at least one of the other groups had to show a reduction in this zero-sum game. Numerically, the loss of White and Asian American students balanced the gains of Mexican American and Latino students, although the percentage loss in each of these groups was small because of the relatively large numbers of White and Asian American students in the sample. The percentage of the eligible African American group that was selected was virtually identical with either model.

Figure 3 divides the Asian American, Mexican American, and Other Latino groups by the language categories from the Student Descriptive Questionnaire. Students who responded “English and another” or “Another” to the question on best language were classified in the English as a second language (ESL) category. For the Asian American ESL students, H(SA) selection resulted in only a slight increase over HVM selection, but in both Latino ESL groups, almost twice as many students were admitted with H(SA) as with HVM. The minimal impact on Asian Americans compared to the Latino groups may be explained by differential test-taking patterns; over 40 percent of the students in the Latino groups took an SAT II Spanish Test, but only 8 percent of the Asian Americans took an Asian language test.

When language tests were excluded from the Subject

Test average, the increase in the number of the Mexican American and Other Latino groups essentially disappeared; the small apparent increase remaining was not statistically significant (standard errors of 1.6 and 2.2 respectively for the percentages in the Mexican American and Other Latino groups). The impact of including or excluding the language tests is somewhat muted because only 43 percent of the Mexican American students and 51 percent of the Other Latino students took one of the Spanish Subject Tests (either Spanish or Spanish with Listening). In order to gauge the impact of the language test on the likelihood of selection, we examined the sample of Mexican American and Other Latino students who had taken one of the Spanish Tests. As shown in Figure 4, in both groups almost twice as many students were selected with the index including the Subject Test average as by the index that used V and M scores. Excluding the Spanish Test from the Subject Test average markedly reduced the number of students selected from these groups. Recall that roughly half of the weight in the prediction equation is on the high school average and, because most students take three Subject Tests, the Spanish Test is approximately one-third of the Subject Test weight (or $\frac{1}{6}$ of the total weight); given this relatively small weight, the effect of including or

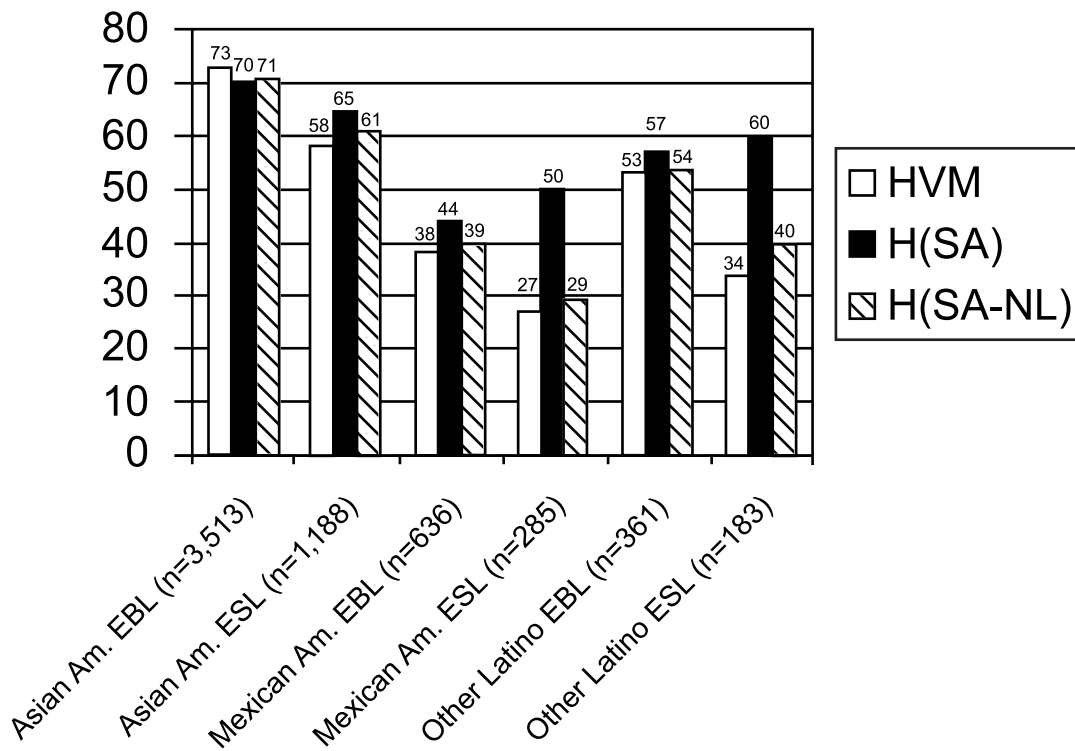


Figure 3. Percent of each ethnic/ESL subgroup selected by HVM, H(SA), and H(SA-NL) for upper $\frac{2}{3}$.

excluding the Spanish Test is indeed dramatic. Test means show the reasons for this relative advantage. In the combined Hispanic groups, the mean score on the Spanish Tests (combining the tests with and without listening) was 147 points higher than the mean score on SAT I: verbal (666 vs. 519; SDs 90 and 91, respectively); in the White sample, the mean score on the Spanish Tests was 85 points lower than the mean score on SAT I: verbal (556 vs. 641; SDs 89 and 77, respectively).

For this group of students who took a Spanish Test, Figure 4 indicates that the selection index that included the average of the Subject Tests resulted in the selection of more Hispanic students than selections based on HVM. We next determined how successful these students were, defining success as achieving a freshman GPA of 2.5 or better, and again using the sample of students who had taken one of the Spanish Subject Tests. For the sample of Mexican American students, including those who were not selected with any of the indices, 59 percent were successful by this criterion. In the Other Latino sample, the overall success rate was 69 percent for the 2.5 or better criterion. As indicated in Figure 5, the students selected by HVM were most successful on a percentage basis; 79 percent of the Mexican American students selected by HVM were successful compared to 66 percent for H(SA). For the

Other Latino students, 84 percent selected by HVM were successful compared to 76 percent for H(SA). If maximizing the percent of successful students in the Hispanic groups were the goal, selections should be based on HVM. However, recall that many more Hispanic students were selected with H(SA) than with HVM. If emphasis is placed on the number of successful students selected from the subgroup instead of on the percent of students in the selected subgroup who are successful, a different conclusion is reached. As indicated in Figure 6, the number of successful Hispanic students was greatest for selections based on the index that used the average of the Subject Tests, including the Spanish Subject Test. If admitting the maximum number of potentially successful Hispanic students were the goal, selections should be based on H(SA).

Selecting the Top One-third

The above analyses assumed that, within each institution, two-thirds of the class would be selected. The following analyses were based on selecting the top one-third within each institution. For the three primary selection models (HVM, H[SA], and H[SA-NL]), the proportion of women selected was the same, 31 percent. Because the pool contained slightly more women than men (7,610 to

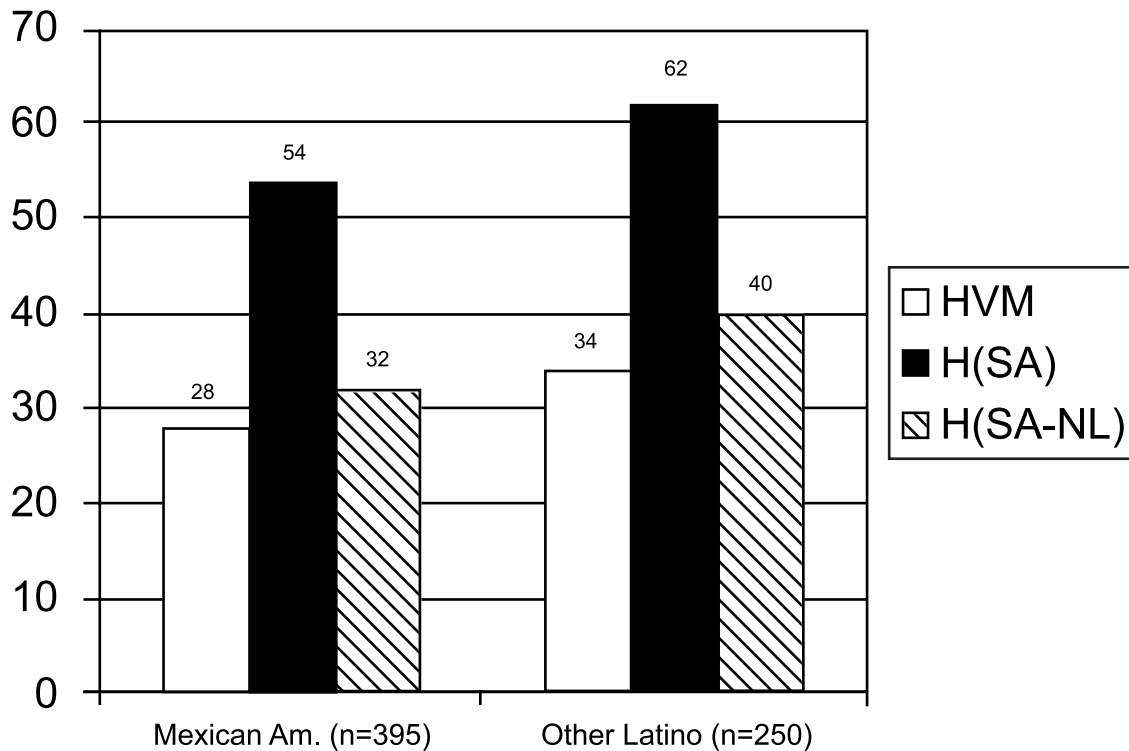


Figure 4. For students who took a Spanish Subject Test, percent of each group selected by HVM, H(SA), and H(SA-NL) for upper $\frac{2}{3}$.

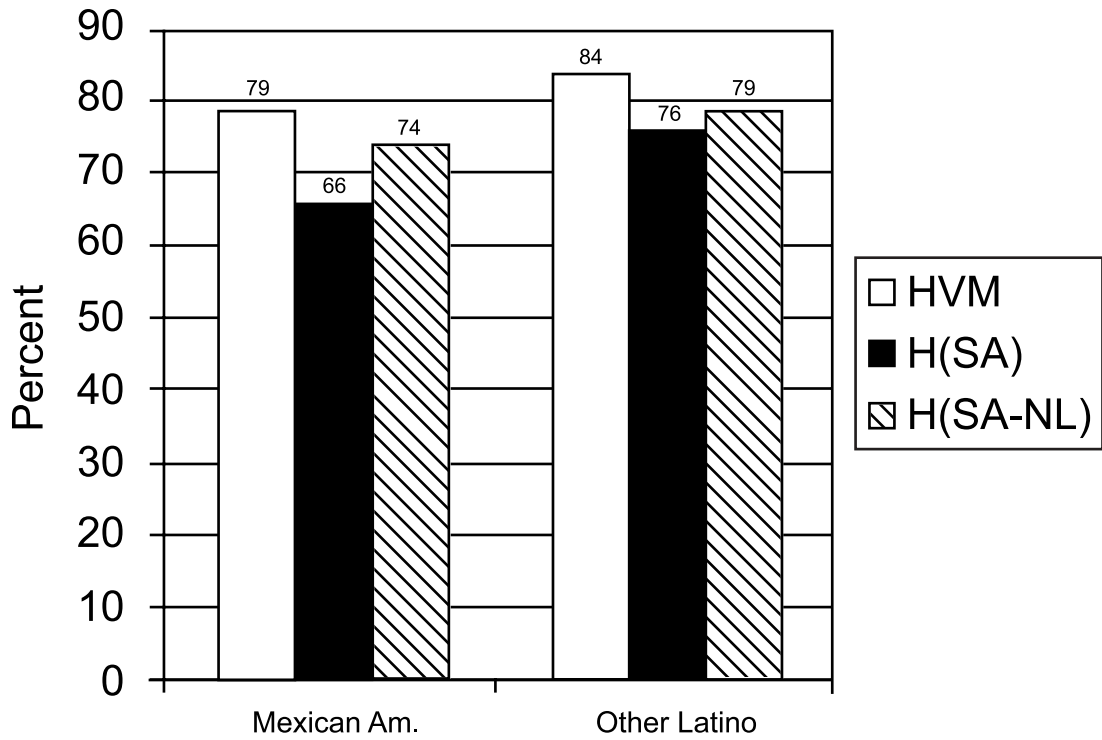


Figure 5. For students who took a Spanish Subject Test, percent of selected students in each group who are successful (GPA 2.5 or higher).

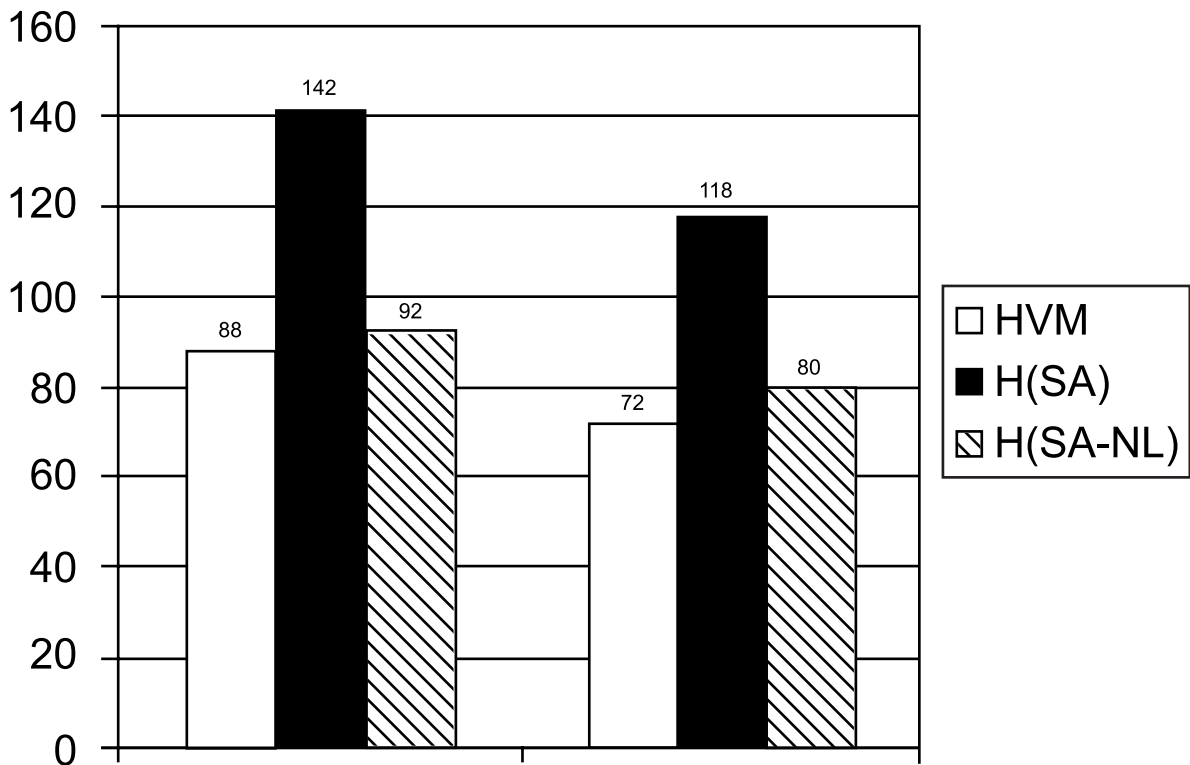


Figure 6. For students who took a Spanish Subject Test, number of selected students in each group who are successful (GPA 2.5 or higher).

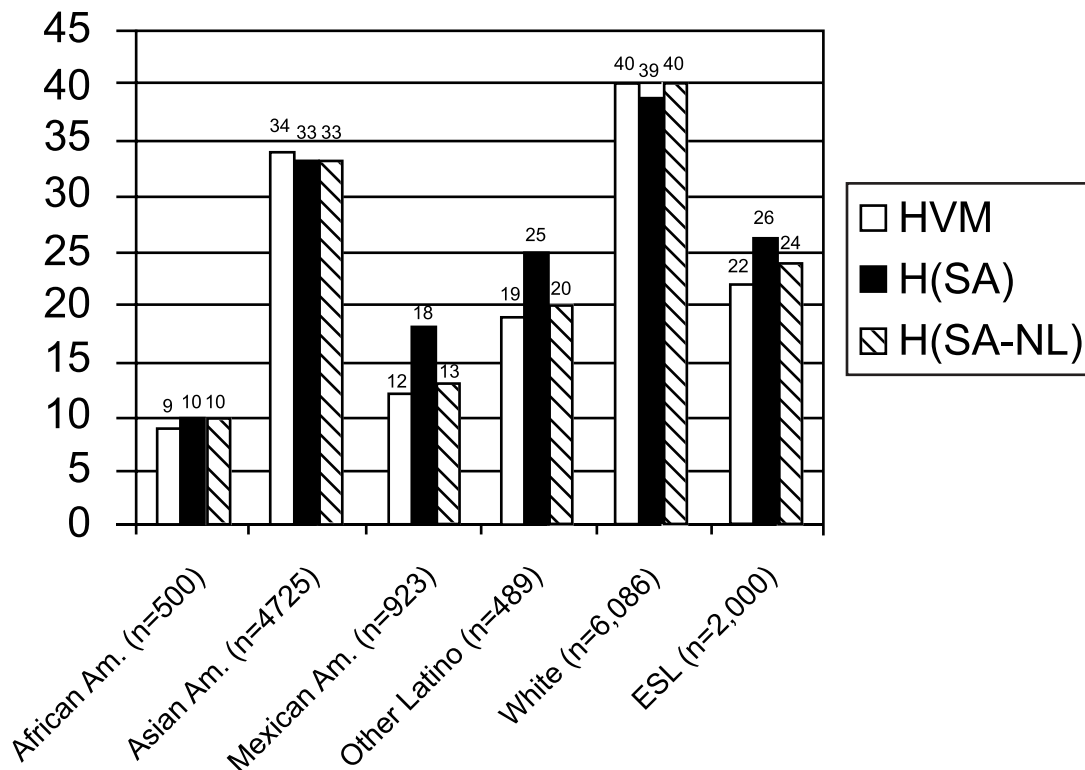


Figure 7. Percent of each group selected by HVM, H(SA), and H(SA-NL) for upper 1/3.

6,246), the number of women selected was almost the same as the number of men selected (for H[SA], 2,376 women and 2,284 men were selected). As with the top two-third selection, adding SAT II: Writing Test to HVM yielded an increase of about 2 percentage points to the percentage of women selected (from 30.7 percent to 32.6 percent), though 93 percent of the selections are the same with HVM as with HVMW. Figure 7 shows the percent selected from each ethnic group for each of the three major indices. The general pattern is the same as was observed for the top two-third selections; for all of the

selection models, White students were overrepresented, Asian American students were proportionally represented, and the other groups were underrepresented relative to their numbers in the applicant population. Mexican American, Other Latino, and ESL students were somewhat more likely to be admitted with the model that used the average of the Subject Tests than with the model that used SAT I V and M scores.

Success rates, once again defining success as achieving a grade point average of at least 2.5, were comparable across the different selection models. Table

TABLE 2

Success (GPA 2.5 or greater) Rates for Students Selected or Rejected in Top 1/3 by HVM and/or H(SA)

| Group | Group 1 (in both) | | Group 2 (out both) | | Group 3 (in H[SA] only) | | Group 4 (in HVM only) | |
|--------------|-------------------|------------|--------------------|------------|-------------------------|------------|-----------------------|------------|
| | n | % GPA 2.5+ | n | % GPA 2.5+ | n | % GPA 2.5+ | n | % GPA 2.5+ |
| Men | 1,869 | 90 | 3,495 | 69 | 415 | 85 | 485 | 84 |
| Women | 1,860 | 95 | 4,755 | 77 | 516 | 88 | 479 | 88 |
| African Am. | 34 | 94 | 437 | 70 | 16 | 88 | 13 | 85 |
| Asian Am. | 1,239 | 91 | 12,816 | 69 | 317 | 87 | 353 | 81 |
| Mexican Am. | 86 | 83 | 781 | 61 | 30 | 73 | 26 | 88 |
| Other Latino | 77 | 96 | 415 | 67 | 32 | 78 | 25 | 88 |
| White | 1,983 | 93 | 3,246 | 81 | 377 | 89 | 480 | 88 |
| Total | 3,729 | 92 | 8,250 | 74 | 931 | 87 | 964 | 86 |

2 shows the number of students selected by both methods, rejected by both methods, and selected by one but rejected by the other. For each of these groups, the percent of the selected students who were successful is also shown. Success rates for students in Group 3 (selected by H[SA] and not by HVM) were virtually identical to success rates in Group 4 (selected by HVM but not H[SA]), except in the two Hispanic groups in which success rates were higher for the HVM selections. The relatively low success percentages in Group 2 (rejected by both HVM and H[SA]) is evidence for the validity of selections based on high school average and either SAT I or SAT II test scores.

Conclusion

Colleges that are selecting students from applicant pools that are similar to the enrolled students in this study could select a class with comparable freshman grades whether they used the SAT II: Subject Tests or the SAT I: Reasoning Test. Switching to the SAT II test average would have a minimal impact on the number of women or African American students selected. Noticeably more Mexican American and Other Latino students would be selected with the Subject Test average, especially if students could submit the Spanish or Spanish with Listening Subject Tests. Adding the SAT II: Writing Test to the SAT I: Reasoning Test would increase the proportion of women selected, but by less than 3 percentage points.

All of the institutions in the current sample were at least moderately selective, and most were highly selective. Further study is needed before generalizations to less selective institutions and more diverse applicant pools could be made. Finally, it is important to recognize that we have modeled only one type of information that goes into complex admission decisions. As noted by Bowen and Bok (1998), "Talk of basing admissions decisions strictly on test scores and grades assumes a model of admissions radically different from the one that exists today" (p. 29).

References

- Barrett, G. V., & Depinet, R. L. (1991). A reconsideration of testing for competence rather than for intelligence. *American Psychologist* 46, 1012–1024.
- Bowen, W. G., & Bok, D. (1998). *The shape of the river: Long-term consequences of considering race in college and university admissions*. Princeton, NJ: Princeton University Press.
- Bridgeman, B., McCamley-Jenkins, L., & Ervin, N. (2000). *Predictions of freshman grade point average from the revised and recentered SAT I: Reasoning Test*. (College Board Report No. 2000-1). New York: College Entrance Examination Board.
- College Board (1999a). *Taking the SAT I: Reasoning Test*. New York: College Entrance Examination Board.
- College Board (1999b). *Taking the SAT II: Subject Tests*. New York: College Entrance Examination Board.
- Crouse, J., & Trusheim, D. (1988). *The case against the SAT*. Chicago: University of Chicago Press.
- Kane, T. J. (1998). Racial and ethnic preferences in college admissions. In C. Jenks and M. Phillips, Eds., *The black-white test score gap*. Washington, D.C.: Brookings Institution Press.
- Linn, R. L. (1994). Performance assessment: Policy promises and technical measurement standards. *Educational Researcher* 23(9), 4–14.
- McClelland, D. C. (1973). Testing for competence rather than for "intelligence." *American Psychologist* 28, 1–14.
- McClelland, D. C. (1994). The knowledge-testing-educational complex strikes back. *American Psychologist* 49, 66–69.
- Messick, S. (1989). Validity. In R.L. Linn (Ed.) *Educational Measurement* (3rd ed., pp. 13–103). New York: American Council on Education and Macmillan.
- Resnick, L. B., & Resnick, D. P. (1992). Assessing the thinking curriculum: New tools for educational reform. In B. R. Gifford & M. C. O'Connor (Eds.), *Changing assessments: Alternative views of aptitude, achievement, and instruction* (pp. 37–75). Boston: Kluwer Academic Publishers.
- Shepard, L. A. (1992). Commentary: What policy makers who mandate tests should know about the new psychology of intellectual ability and learning. In B. R. Gifford & M. C. O'Connor (Eds.), *Changing assessments: Alternative views of aptitude, achievement, and instruction* (pp. 301–328). Boston: Kluwer Academic Publishers.
- Shepard, L. A. (1997). The centrality of test use and consequences for test validity. *Educational Measurement: Issues and Practice* 16(2), 5–8, 13, 24.
- UCSD (2000). UCSD selection criteria. (http://admissions.ucsd.edu/ad_fresh2.html).
- Wightman, L. F. (1997). The threat to diversity in legal education: An empirical analysis of the consequences of abandoning race as a factor in law school admission decisions. *New York University Law Review* 72(1), 1–53.
- Willingham, W. W., & Breland, H. M. (1982). *Personal qualities and college admissions*. New York: College Entrance Examination Board.
- Wing, C. W., & Wallach, M. A. (1971). *College admissions and the psychology of talent*. New York: Holt, Rinehart and Winston.

