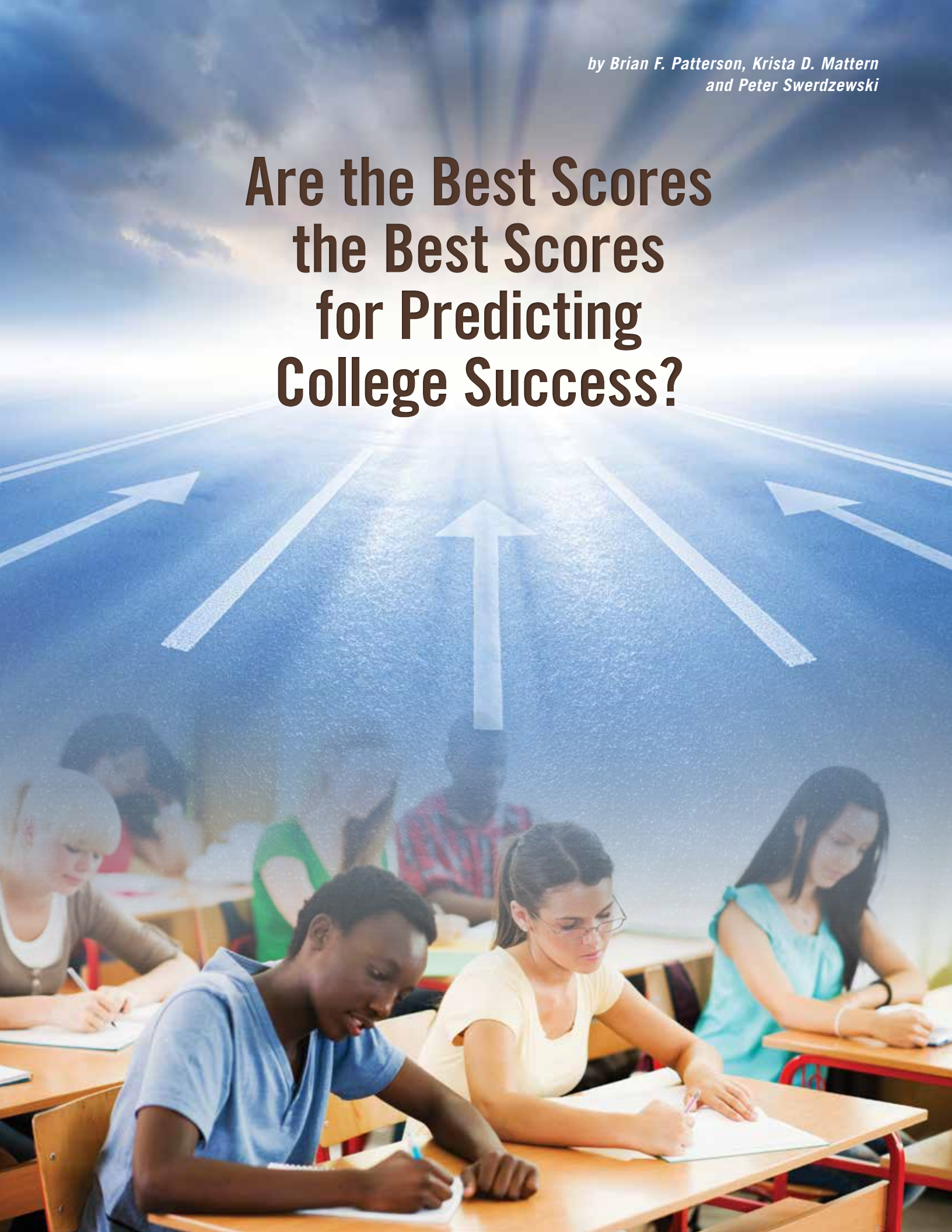
by Brian F. Patterson, Krista D. Mattern and Peter Swerdzewski

Are the Best Scores the Best Scores for Predicting College Success?



Abstract

The College Board's SAT® Score Choice™ policy allows students to choose which set(s) of scores to send to colleges and universities to which they plan to apply. Based on data gathered before the implementation of that policy, the following study evaluated the predictive validity of the various sets of SAT scores. The value of five score sets for predicting first-year college grade-point average (FYGPA) was examined using a sample of 150,377 students from 110 institutions. These sets of scores were from students': (1) first administration, (2) latest administration, (3) section averages, (4) highest single administration, and (5) highest individual sections. The various score sets were found to have nearly equal predictive validity for the total group, by various institutional characteristics, and regardless of how often students tested.

Background on the Use of Tests in College Admission

Scores from standardized tests are a critical component of the admission process at nearly all competitive colleges and universities (i.e., institutions). According to the 2010 NACAC *State of College Admission* report, 89.8 percent of colleges attributed either "considerable" or "moderate" importance to admission test scores in the admission process. The top factors in selecting applicants—in order—were: (1) grades in college preparatory courses; (2) strength of curriculum; (3) standardized admission test scores; and (4) overall high school grade point average (Clinedinst and Hawkins, 2010).

Students understand the important role that standardized test scores play in the admission process: to increase the likelihood of being admitted to institutions with competitive admission processes, students must have strong standardized test scores and a successful high school profile. With three SAT scores per administration [i.e., critical reading (SAT-CR), mathematics (SAT-M), and writing (SAT-W)], those students who take the SAT numerous times could have upwards of a dozen scores. With so many scores available to admission officers who are tasked with creating the most competent entering class possible, an in-depth study of how best to combine scores to predict student performance is necessary.

Sifting Among the Score Sets

Students interested in presenting a competitive admission portfolio frequently take a standardized test numerous times. For example, more than half of the SAT-takers graduating high school in 2006—referred to as the 2006 College-Bound Seniors and including more than 1.3 million students—took the SAT more than once (see Table 1). In the past, when admission officers reviewed applications from these students, they had access to all standardized scores from each applicant; however, admission officers within a given university in compliance with NACAC's goals of equity are likely to use only one set of test scores when making admission decisions. In other words, they are likely to have made a policy decision to uniformly use, for example, the average of all admission test scores submitted by applicants. Examples of how to fairly treat applicants who took the SAT more than once include:

- *The first set of scores:* This score set is simply the SAT-CR, SAT-M and SAT-W section scores from the first time the applicant took the SAT.
- *The latest set of scores:* This policy implies using the SAT-CR, SAT-M and SAT-W section scores from the latest (i.e., most recent) instance in which the applicant took the SAT.

…An examinee may now choose whether or not to send scores from a particular test administration. All section scores—critical reading, mathematics, and writing—from a given test date must be sent or withheld…

- *The within-section average scores:* Admission officers using this policy take the average (i.e., mean) of each section score across all SAT administrations (i.e., testing occasions). For example, if a student took the SAT in March and May, the average of his or her two SAT-CR section scores, the average of the two SAT-M scores and the average of the two SAT-W scores (i.e., three separate within-section averages) would be used in the admission decision.

- *The highest single administration:* This score set represents the SAT-CR, SAT-M and SAT-W section scores associated with the highest composite score (i.e., the sum of the scores from the three sections) from a single administration. That is, even if an applicant's highest SAT-CR score is from the first time she took the SAT and her highest SAT-M and SAT-W scores are from the second time she took the SAT, only the three section scores from a single administration that add up to her highest single administration score are used in the admission process. This method is often called the "highest combined" or "best administration" approach and is called the "Highest Sitting" approach by The College Board in reference to the SAT Score Choice™ policy (College Board, 2010a).[1]

- *The highest individual sections:* As opposed to the highest single administration, admission officers using this policy select the applicant's highest SAT-CR, SAT-M and SAT-W section scores across all test administrations. That is, if an applicant's highest SAT-CR score was from the first time he took the SAT, and his highest SAT-M and SAT-W scores were from the second time he took the SAT, his first SAT-CR score and second SAT-M and SAT-W scores would be used in the admission process. This method is often called the "highest individual," "cherry picking" approach, or the "Highest Section" approach in the context of the SAT Score Choice policy (College Board, 2010a) and always results in the highest composite score for use in admission decisions.

Rigorous research on the prevalence of each score set policy—in other words, the institution-specific policy on how to use multiple admission test scores in the selection process—does not yet exist and is not the subject of this study. Anecdotally the most common score set policy appears to be the highest individual sections (College Board, 2010a; College Board, 2009).[2] The College Board, on the other hand, bases much of its research on examinees' latest scores, because performance on the latest test administration is expected to have the strongest relationship with the criteria (e.g., first-year college performance) being studied. This hypothesis is due to the simple fact that the latest test scores are closer in time to when the criterion is measured. The least common approaches used by colleges and universities appear to be either to use each applicant's first set of scores, the latest set of scores, or to use examinees' average score across all administrations, but all five of these score sets are considered in this study for completeness.

SAT Score Choice™: Another Element Relevant to Selecting the Best Scores

Historically, when an examinee requested that The College Board send his or her official SAT scores to college admission officers, The College Board has sent all of the examinee's scores. That is, The College Board forwarded each section score from every SAT completed by the examinee to the requested institution(s). Beginning in the spring of 2009, examinees began having the option to specify which scores should be sent to which institutions. Specifically, an examinee may now choose whether or not to send scores from a particular test administration. All section scores—critical reading, mathematics, and writing—from a given test date must be sent or withheld; it is not possible to send certain section scores from one administration and other section scores from another administration. For example, an examinee who took the SAT in March and again in May can send his or her critical reading, mathematics and writing scores from the March administration of the SAT and choose not to send the three section scores from the May administration. This policy (SAT Score Choice) was designed to reduce the stress associated with

high-stakes standardized testing and to improve the test-day experience. For more information on the SAT Score Choice policy, see www.collegeboard.com/scorechoice.

Because of this policy, admission officers only have available to them those scores that applicants choose to send, unless they require applicants to send all scores. For those applicants who do send all scores, current admission policies—whether they use the average, highest individual sections or other approach—may be used and will have the same effect as before the implementation of SAT Score Choice. However, for those examinees who take advantage of the SAT Score Choice policy, admission officers need to rely on the scores that students choose to send.

The Need to Understand Which Set of Scores is the Best Set of Scores

In the past, admission officers have discussed which set of scores should be used in making admission decisions (e.g., Whetstone, 2008); however, neither standards nor recommendations exist that explicitly state which set of scores should be used in the admission process. Although the NACAC Statement of Principles of Good Practice (NACAC, 2010), the Standards for Educational and Psychological Testing (AERA, APA and NCME, 1999) and The College Board's Guidelines on the Use of College Board Test Scores and Related Data (The College Board, 2010b) are all silent on the matter, all three publications implicitly or explicitly indicate that college and university admission officers are obligated to ensure that their admission processes are rigorous and fair. This study is aimed at assisting admission officers in this endeavor by providing research on which set of SAT scores was most predictive of college success. Specifically, the authors address the following questions in this study:

Question #1: How often would the use of different score set policies result in different scores considered in admission decisions?

Specifically, what percentage of examinees take the SAT more than once, thereby resulting in more than one set of scores available to admission officers? Furthermore, it is important to recognize that, although many students take the SAT more than once, students' scores do not always differ across administrations in such a way that a score set policy would make a difference. Therefore, it was informative to evaluate the prevalence of instances in which different score set policies would result in different composite scores for consideration by admission professionals. In other words, how often do examinees have identical scores regardless of the score set used, thereby making irrelevant the policy that specifies the set of scores to be used to select students for admission to college?

Question #2: Which set of scores are most predictive of college success?

To answer this question, students' first-year college grade-point average (FYGPA) was used as a measure of student success in college. If there were wide variation in the predictive validity of SAT by score set policy, then it would be even more imperative that admission professionals carefully consider their policies.

Question #3: Do institutional characteristics or student test-taking frequency affect the degree to which various sets of scores predict FYGPA?

Specifically, are the different sets of scores more or less predictive of FYGPA across institutions of differing size, selectivity or control (i.e., public or private)? Are the various sets of scores more or less predictive for students who take the SAT once, twice, three times, or four or more times? These are all important considerations in the choice of which score set policy to put into place.

For those applicants who do send all scores, current admission policies—whether they use the average, highest individual sections or other approach—may be used and will have the same effect as before the implementation of SAT Score Choice. However, for those examinees who take advantage of the SAT Score Choice policy, admission officers need to rely on the scores that students choose to send.

Methods

Sample

A sample of 150,377 first-time postsecondary students who graduated from high school in 2006 and entered college in the fall of 2006 was used in this study. All students took the current version of the SAT at least once, reported their high school grade point average (HSGPA), and subsequently enrolled in one of 110 postsecondary institutions that participated in The College Board's

National SAT Validity Study (Kobrin, Patterson, Shaw, Mattern, and Barbuti, 2008). These 110 institutions are adequately representative of the population of four-year institutions in the United States (for additional information on the sample used in the current study, see Kobrin et al., 2008). SAT scores were obtained from College Board databases and students' FYGPAs were provided by the 110 institutions in the sample.

Measures

The SAT consists of critical reading (SAT-CR), mathematics (SAT-M) and writing (SAT-W) sections, each of which ranges from 200 to 800 by 10-point increments. Along with the HSGPA—whose values range from 0.00 to 4.33—provided on the SAT Questionnaire during the registration process, the SAT is widely used in undergraduate college admission. One of the most common outcomes when analyzing the validity of college admission tests is students' first-year grade point average (FYGPA) in college. That variable had a fairly consistent range of 0.00 to 4.00, but some institutions' grade scales did go as high as 4.3.

Data Analysis

To determine the frequency with which different score set policies would result in different information (i.e., conflicting test scores), the number of students who take the SAT once, twice, three times, and four or more times was first computed. For students who took the SAT at least twice, the percentage of students who would have exactly the same composite scores (the sum of the critical reading, mathematics and writing sections) across all potential pairs of the five score sets was computed. For example, if a student received a 500 on each of the three sections during his or her latest administration of the SAT, and these three scores are also the highest scores for this student across all occasions in which the student took the SAT, this student's composite score from his or her latest test administration and this student's composite score from his or her highest three individual sections are equal (i.e., both sum to 1500).

Similar to previous research examining the validity of the SAT for predicting FYGPA (e.g., Agronow and Studley, 2007; Kobrin, et al., 2008; Mattern, Patterson, Shaw, Kobrin, and Barbuti, 2008; Cornwell, Mustard, and Van Parys, 2008; Patterson, Mattern, and Kobrin, 2009; Patterson and Mattern, 2011; Patterson and Mattern, 2012), the multiple correlations of different SAT score sets and HSGPA with FYGPA were computed. Multiple correlations, like simple (i.e., bivariate) correlations, describe the strength of the linear relationship among variables. Correlations less than -.5 or more than +.5 are commonly thought of as indicating a strong linear relationship among variables (Cohen, 1988). Both raw and restriction-of-range corrected correlations based on the Pearson-Lawley multivariate correction method are reported (Lawley, 1943). To reflect the differences in predictive validity across institutions, the various sets of SAT scores and FYGPAs were correlated within each institution and then pooled across institutions by weighting each correlation by the sample size (i.e., number of students who contributed to the estimation of the respective correlation) in a method similar to what was used by Powers (2004).

Results

Question #1: How often would the use of a particular score set policy result in different scores used in admission decisions?

Table 1 shows that 62 percent of students in the sample took the SAT more than once. Comparing those students with the full cohort of 2006 College-Bound Seniors, the examinees in this study's sample were more likely to re-test than those in the national cohort (62 percent in the sample vs. 47 percent in the population). It is important to consider that the sample differs from the national cohort in that the national cohort includes students who: (a) did not enroll at a postsecondary institution at all; (b) delayed their enrollment; and (c) enrolled at two-year community colleges.

Table 1. Frequency with Which Students Took the SAT

Number of SAT Administrations	Current Sample		2006 College-Bound Seniors[a]	
	N	%	N	%
1	57,743	38	729,326	53
2	73,691	49	520,068	38
3	17,192	11	113,748	8
4+	1,751	1	13,598	1
2+	92,634	62	647,414	47
Total	150,377	100	1,376,740	100

[a]: Limited to the 2006 College-Bound Seniors who reported their high school grade point average. For more information on the 2006 College-Bound Seniors Cohort, see College Board (2006). Percentages may not sum to 100 percent due to rounding.

As displayed in Table 2, the composition and characteristics of the sample reveal a few patterns worthy of note. First, the students in this sample took the SAT, on average, 1.75 times, with females testing slightly more often than males (1.79 and 1.72 times, respectively). Asian, Asian-American or Pacific Islander students tended to test more often (1.99 times) and American Indian or Alaska Native students, on average, tested less often (1.66 times) than students of other racial/ethnic identities. In terms of students' best spoken language, those who indicated English only as their best tested the least often (1.74 times), while those students whose best language was a

Table 2. Descriptive Statistics by Student Characteristics

		Mean (Standard Deviation)							
					SAT Composite (CR + M + W)				
	N	Number of SAT Admins	HSGPA	FYGPA	First	Latest	Average	Highest Single Admin.	Highest Individual Sections
Overall	150,377	1.75 (0.70)	3.60 (0.50)	2.97 (0.71)	1668 (255)	1693 (255)	1681 (252)	1707 (252)	1721 (252)
Gender									
Female	81,039	1.79 (0.70)	3.65 (0.48)	3.05 (0.67)	1648 (253)	1672 (254)	1660 (251)	1687 (251)	1702 (251)
Male	69,338	1.72 (0.70)	3.55 (0.52)	2.88 (0.74)	1693 (255)	1716 (255)	1705 (252)	1730 (251)	1744 (251)
Racial / Ethnic Identity									
American Indian or Alaska Native	794	1.66 (0.68)	3.53 (0.54)	2.78 (0.76)	1604 (234)	1629 (233)	1616 (230)	1642 (232)	1654 (232)
Asian, Asian-American, or Pacific Islander	14,058	1.99 (0.76)	3.66 (0.47)	3.05 (0.67)	1711 (272)	1748 (269)	1730 (267)	1765 (266)	1783 (265)
Black or African-American	10,243	1.78 (0.71)	3.40 (0.55)	2.63 (0.74)	1485 (230)	1508 (232)	1496 (227)	1524 (229)	1539 (229)
Hispanic, Latino, or Latin American	10,599	1.72 (0.69)	3.59 (0.51)	2.73 (0.77)	1558 (242)	1581 (247)	1569 (242)	1595 (245)	1608 (246)
White	103,507	1.73 (0.69)	3.62 (0.49)	3.02 (0.69)	1689 (243)	1712 (244)	1700 (241)	1726 (240)	1739 (240)
Other	4,476	1.79 (0.69)	3.57 (0.50)	2.95 (0.71)	1658 (260)	1683 (264)	1671 (259)	1699 (259)	1714 (259)
No Response	6,700	1.72 (0.68)	3.63 (0.50)	3.05 (0.69)	1730 (270)	1755 (271)	1743 (267)	1769 (266)	1782 (266)
Best Language									
English Only	139,767	1.74 (0.70)	3.60 (0.50)	2.98 (0.71)	1675 (252)	1698 (252)	1686 (249)	1712 (249)	1726 (249)
English and Another Language	7,361	1.91 (0.75)	3.61 (0.49)	2.90 (0.72)	1602 (274)	1635 (277)	1619 (272)	1651 (274)	1668 (275)
Another Language	1,678	2.04 (0.80)	3.61 (0.52)	3.04 (0.68)	1495 (264)	1547 (270)	1522 (263)	1563 (269)	1582 (271)
No Response	1,571	1.71 (0.71)	3.53 (0.54)	2.88 (0.74)	1618 (292)	1640 (295)	1629 (291)	1655 (290)	1668 (290)
Number of SAT Administrations									
1	57,743	1.00 (0.00)	3.64 (0.51)	2.96 (0.76)	1709 (270)	1709 (270)	1709 (270)	1709 (270)	1709 (270)
2	73,691	2.00 (0.00)	3.59 (0.50)	2.99 (0.68)	1655 (243)	1688 (248)	1672 (241)	1710 (243)	1730 (243)
4+	1,751	4.11 (0.36)	3.56 (0.52)	2.93 (0.65)	1540 (227)	1628 (239)	1588 (226)	1665 (229)	1702 (226)
Note. SAT composite equals the sum of SAT sections (critical reading + mathematics + writing; CR + M + W).									

language other than English tested most often (2.04 times). Students who tested only once had higher HSGPA (M = 3.64) than students who tested multiple times (M = 3.58). The phenomenon of lower first SAT scores being associated with students who tested more often may be initially counterintuitive, but it has been documented in prior research (e.g., Alderman, 1981). Therefore, while students' latest SAT scores tend to be higher than their first, some students may choose to take the SAT numerous times with hopes of achieving their target scores with additional testing.

Also worthy of note is that the difference in SAT means among the various score combinations was relatively small. For example, across all students in the sample, the average composite score using the latest score set was 1693, whereas the average composite score using the highest individual sections method was 1721—a mean difference of only 28 points. Comparing the average from the highest individual section scores (M = 1721) to the average from the highest single administration (M = 1707), the difference is even smaller.

Just as differences by student characteristics are of interest, so too are variations by institutional characteristics in this sample. The same statistics displayed in Table 2 are shown in Table 3, but results are summarized by institutional characteristics of control

Table 3. Descriptive Statistics by Institutional Characteristics

		Mean (Standard Deviation)							
					SAT Composite (CR + M + W)				
	N	Number of SAT Admins	HSGPA	FYGPA	First	Latest	Average	Highest Single Admin.	Highest Individual Sections
Overall	150,377	1.75 (0.70)	3.60 (0.50)	2.97 (0.71)	1668 (255)	1693 (255)	1681 (252)	1707 (252)	1721 (252)
Institutional Control									
Public	104,881	1.71 (0.70)	3.58 (0.50)	2.91 (0.74)	1631 (239)	1653 (239)	1642 (236)	1667 (235)	1680 (235)
Private	45,496	1.86 (0.70)	3.66 (0.49)	3.12 (0.60)	1754 (267)	1784 (268)	1769 (264)	1800 (263)	1816 (262)
Institutional Size									
Small	6,430	1.81 (0.69)	3.52 (0.53)	2.94 (0.69)	1641 (264)	1667 (266)	1654 (261)	1683 (261)	1697 (261)
Medium	30,142	1.84 (0.70)	3.57 (0.52)	3.06 (0.66)	1684 (280)	1712 (281)	1698 (277)	1727 (276)	1743 (275)
Large	40,625	1.81 (0.70)	3.52 (0.52)	2.94 (0.70)	1633 (259)	1658 (260)	1645 (256)	1674 (256)	1689 (255)
Very Large	73,180	1.68 (0.69)	3.67 (0.47)	2.96 (0.73)	1684 (238)	1707 (238)	1695 (235)	1719 (236)	1732 (236)
Percentage of Applicants Admitted									
Under 50%	27,110	1.88 (0.71)	3.75 (0.44)	3.20 (0.54)	1837 (256)	1870 (255)	1853 (252)	1884 (249)	1901 (248)
50 to 75%	83,901	1.77 (0.70)	3.60 (0.49)	2.96 (0.70)	1652 (239)	1676 (238)	1664 (235)	1691 (234)	1705 (233)
Over 75%	39,366	1.63 (0.68)	3.52 (0.53)	2.84 (0.77)	1587 (233)	1606 (233)	1597 (230)	1619 (230)	1630 (230)

Note. Institution sizes (number of undergraduates) are as follows: small: 750 to 1,999; medium: 2,000 to 7,499; large: 7,500 to 14,999; and very large: 15,000 or more. SAT composite equals the sum of SAT sections (critical reading + mathematics + writing; CR + M + W).

(i.e., public or private), size (i.e., total degree-seeking undergraduates), and selectivity (as measured by undergraduate admittance rate). Students attending private institutions tended to take the SAT more often (1.86 times) than those attending public institutions (1.71 times). Those attending very large institutions tended to test less often (1.68 times) than students at smaller (i.e., small, medium and large) institutions (1.82 times). And finally, as institutional admittance rate decreases—that is, as selectivity increases—students tended to take the SAT more often, with those attending schools that admit over 75 percent of applicants taking the SAT on average 1.63 times and those attending the most selective group of institutions taking the SAT 1.88 times.

If a student takes the SAT only once, there would be only one set of scores available for use by admission officers at institutions to which that student applied. Indeed, for nearly 40 percent of the students whose applications were accepted by the 110 institutions that contributed to this study, only a single set of scores was available for review by admission officers.

A second scenario also exists in which the choice between a pair of score set policies is inconsequential: when a student has the same score regardless of which of the two score sets were used. For example, if a student takes the SAT and receives a 580 each on the critical reading, mathematics and writing sections, then retakes the SAT and receives a 600 on each of the three sections, his or her latest scores, highest single administration scores, and highest individual sections are the same: 600s across all three sections, for a composite score of 1800. In other words, when a college considers whether to use score set policy A or B, they ought to consider that there may be some substantial group of applicants for whom admission test scores under either policy would be equal.

Given the maturation of students that is expected to occur over the latter half of high school, one may expect that a large number of examinees would receive their best set of scores on their final (latest) administration of the assessment. The evidence presented as Table 4 generally supports this hypothesis with regard to the highest single administration and latest score sets. Of the 92,634 students who took the SAT more than once, 64 percent of examinees achieved their highest single administration score on their latest SAT. For these students, the information available to admission professionals for decision-making would be identical under either the policy of using the latest SAT or the highest single administration of the SAT.

About one in three examinees achieve their highest individual section scores during a single administration, thus their highest single administration and highest individual section score sets are equal. Moreover, one quarter of students demonstrated their highest level of performance on all three sections of the SAT simultaneously during their latest administration of the assessment. In sum—for a large percentage of the cases and most notably when considering latest and highest single administration—the different score sets would result in the same SAT composite score.

Table 4. Percent of Equal SAT Composites[a] by SAT Score Set for Multiple SAT-takers

	Score Set			
Score Set	**First**	**Latest**	**Average**	**Highest Single Administration**
Latest	4%	-	-	-
Average	8%	8%	-	-
Highest Single Administration	33%	64%	10%	-
Highest Individual Sections	9%	25%	0%	35%

N = 92,634.
[a]: SAT composite equals the sum of SAT sections (critical reading + mathematics + writing; CR + M + W).

Question #2: Which set of scores is most predictive of college success?

Recognizing the need for having a policy on the set of scores to be used in the admission process, it is important to determine which set of scores is the best set of scores to use when predicting students' early college outcomes. In order to do so, the multiple correlations between different combinations of SAT scores and HSGPA with FYGPA were estimated. Note that The College Board recommends that admission officers use the three sections of the SAT in conjunction with HSGPA to predict college success; therefore, multiple correlations between the various SAT score sets, in combination with HSGPA, are reported along with correlation between FYGPA and HSGPA alone and the three SAT sections alone.

Table 5. Corrected (Raw) Correlations[a] of FYGPA with SAT Score Sets and HSGPA

	Corrected (Raw) Correlation with FYGPA		
Score Set	**HSGPA**	**CR, M, W**	**HSGPA, CR, M, W**
First	0.54 (0.37)	0.52 (0.34)	0.61 (0.44)
Latest	0.54 (0.37)	0.54 (0.35)	0.62 (0.45)
Average	0.54 (0.37)	0.54 (0.36)	0.62 (0.45)
Highest Single Administration	0.55 (0.37)	0.54 (0.35)	0.62 (0.45)
Highest Individual	0.55 (0.37)	0.54 (0.35)	0.63 (0.45)

Note: N = 150,377. SAT critical reading, mathematics and writing sections abbreviated CR, M and W, respectively. Institution sizes (number of undergraduates) are as follows: small: 750 to 1,999; medium: 2,000 to 7,499; large: 7,500 to 14,999; and very large: 15,000 or more.
[a]: Corrected and raw correlations represent the multiple correlation between each set of predictors with students' FYGPAs. Corrected correlations have been corrected for restriction-of-range and pooled across each of the 110 institutions. Slight variation in the corrected correlation between high school and FYGPA are due to the fact that the correction is multivariate in nature and not due to sample differences.

Among the sample of students examined in this study, the predictive validity of the various sets of scores when used alone or in conjunction with HSGPA is quite high for all score sets and likewise quite similar across score sets. Table 5 shows that the corrected multiple correlations of FYGPA with HSGPA and SAT scores ranged from r* = .61 to .63 across the five score sets (corresponding raw multiple correlations ranged from r = .44 to .45; for restriction-of-range corrected and raw multiple correlations for additional combinations of predictors, such as SAT critical reading and mathematics, see Appendices B through F, available by request from the authors). If HSGPA were not included, the corrected multiple correlations of FYGPA with SAT scores ranged from r* = .52 to .54 across the five score sets (corresponding raw multiple correlations ranged from r = .34 to .36). Since HSGPA reflects unique and important aspects of academic preparedness, the multiple correlations presented and discussed throughout the remainder of the study generally refer to the multiple correlation of FYGPA with HSGPA and the separate SAT section scores (critical reading, mathematics and writing).

If a student takes the SAT only once, there would be only one set of scores available for use by admission officers at institutions to which that student applied. Indeed, for nearly 40 percent of the students whose applications were accepted by the 110 institutions that contributed to this study, only a single set of scores was available for review by admission officers.

The two score sets that are quite prevalent for use in undergraduate admission—highest single administration scores and highest individual sections—have corrected multiple correlations with FYGPA that do not substantially differ. Students' first scores have the lowest predictive power; however, it must be reiterated that the predictive validity of students' first scores are not much different from that of the other score sets. In sum, for admission purposes, the five different combinations of predictors have essentially the same predictive validity. Thus, among the five score sets considered in this study, it appears that applicant selection will not be undermined by the decision of which SAT score set an institution employs for admission purposes.

Question #3: Do institutional characteristics or student test-taking frequency affect the degree to which various sets of scores predict FYGPA?

The predictive validity of the various score sets was examined by separate levels of institutional control, institutional size and institutional selectivity. Multiple correlations between the various SAT score sets, in combination with HSGPA, are reported in Table 4. It is also important to note that differences among the student populations that attend each type of institution, rather than differences among the institutions themselves, may be the primary cause of differences in predictive validity across institutions. Raw and restriction-of-range corrected multiple correlations with additional sets of predictors (e.g., SAT critical reading and mathematics scores in combination with HSGPA as predictors of FYGPA) are reported in Appendices B through F, available by request from the authors. For a more in-depth discussion of predictive validity by institutional characteristics, see Kobrin, et al. (2008).

Institutional Control. For private and public institutions, the predictive validity of the various sets of scores is nearly equal. Within public institutions, the various sets of scores, when combined with HSGPA, have multiple correlations of r* = .59 to .61 with FYGPA. The score sets and HSGPA tend to be more predictive at private institutions; the score sets and HSGPA are correlated with FYGPA from r* = .64 to .66. For both public and private institutions in a practical admission context, the various score sets have essentially equal predictive validity, with multiple correlations differing by no more than .02.

Institutional Size. Within each institutional size category (small, medium, large, and very large, determined by the total number of degree-seeking undergraduates), the corrected multiple correla-

They've earned it.
We'll help pay for it.

AXA Equitable, through its philanthropic program, AXA Achievement℠, is proud to be one of the nation's largest corporate providers of scholarships.

AXA Achievement provides both national and local scholarships directly to students in your community.

For more information, please log on to www.axa-achievement.com.

AXA Achievement℠

AXA Achievement℠ is funded by the AXA Foundation, the philanthropic arm of AXA Equitable.

"AXA Achievement℠" is a service mark of the AXA Foundation. The "AXA Scholarship" program is not associated with the National Merit Scholarship Corporation's "Achievement Scholarship" program.

G24947

AXA Equitable Life Insurance Company, NY, NY 10104, (212) 314-4600

> …It must be reiterated that the predictive validity of students' first scores are not much different from that of the other score sets. In sum, for admission purposes, the five different combinations of predictors have essentially the same predictive validity.

tion by SAT score set and HSGPA varied minimally (i.e., < .03). The score sets and HSGPA tended to be more predictive in small institutions (r* = .66 to .68) than very large institutions (r* = .60 to .62), which is consistent with previous findings that the SAT scores are more predictive in smaller institutions (Kobrin, et al., 2008; Patterson, et al., 2009; Patterson and Mattern, 2011; Patterson and Mattern, 2012). As was found with institutional control, the various score sets have similar predictive validity within each institutional size category.

Institutional Selectivity. In line with findings for institutional control and institutional size, the various score sets and HSGPA have approximately equal predictive validity within each category of institutional selectivity. That is, for the most selective institutions in the sample (those that admit fewer than 50 percent of applicants), the least predictive of the score sets, students' first scores, have a corrected multiple correlation of r* = .64 with FYGPA, whereas the most predictive of the score sets, students' highest individual sections, have a corrected multiple correlation that is only .02 greater. Among the least selective institutions in the sample (those that admit more than 75 percent of applicants), the corrected multiple correlations of the various score sets and HSGPA with FYGPA ranged from r* = .60 to .61. In short, the score sets and HSGPA are more predictive in more selective institutions, but the score sets are nearly equally-predictive within each category of selectivity, with correlations differing by at most .02.

Student Test-taking Frequency. To what degree do the various score sets differentially predict students' FYGPAs for those students who take the SAT once, twice, three times, and four or more times? Similar to the predictive validity of score sets for the total sample and across institutional characteristics, the predictive validity of the various score sets is approximately equal within each group of students (see Table 6; further detail is available in Appendices B through F). That is, for those students who took the SAT only once, the corrected multiple correlation is equal (r* = .60) across the various score sets because the same scores are used for the five score sets (i.e., first = latest = average = highest single administration = highest individual sections). For those students who took the SAT four or more times, there is minimal variation in corrected multiple correlations: from r* = .66 to r* = .67. All five score sets plus HSGPA led to similar predictive validity coefficients, with a maximum difference of only .02, for any test-taking frequency considered.

Table 6. Corrected Correlations[a] of FYGPA with SAT Score Sets and HSGPA by Institutional Characteristics and Number of Times Students Took the SAT

	Correlation				
	First	Latest	Average	Highest Single Administration	Highest Individual Sections
Overall	0.61	0.62	0.62	0.62	0.63
Institutional Control					
Public	0.59	0.61	0.60	0.61	0.61
Private	0.64	0.65	0.65	0.66	0.66
Institutional Size					
Small	0.66	0.67	0.67	0.68	0.68
Medium	0.62	0.63	0.63	0.64	0.64
Large	0.60	0.62	0.62	0.62	0.63
Very Large	0.60	0.61	0.61	0.61	0.62
Percentage of Applicants Admitted					
Under 50%	0.64	0.66	0.66	0.66	0.66
50 to 75%	0.60	0.61	0.61	0.62	0.62
Over 75%	0.60	0.60	0.60	0.61	0.61
Number of SAT Administrations					
1	0.60	0.60	0.60	0.60	0.60
2	0.62	0.63	0.64	0.63	0.64
3	0.61	0.63	0.64	0.64	0.64
4+	0.66	0.66	0.67	0.66	0.67

Note: N = 150,377. Multiple correlations include SAT section scores and HSGPA as predictors. Institution sizes (number of undergraduates) are as follows: small: 750 to 1,999; medium: 2,000 to 7,499; large: 7,500 to 14,999; and very large: 15,000 or more.
[a]: Correlations are corrected for restriction-of-range and pooled across each of the 110 institutions.

Conclusion and Implications for the Admission Process

In general, it is understood that institutions need a policy that specifies the set of scores that is to be considered in the admission process because nearly half of all examinees take the SAT more than once and thus provide more than one set of scores. While many students test more than once, it must be noted that few tended to take the SAT four or more times. The implication of

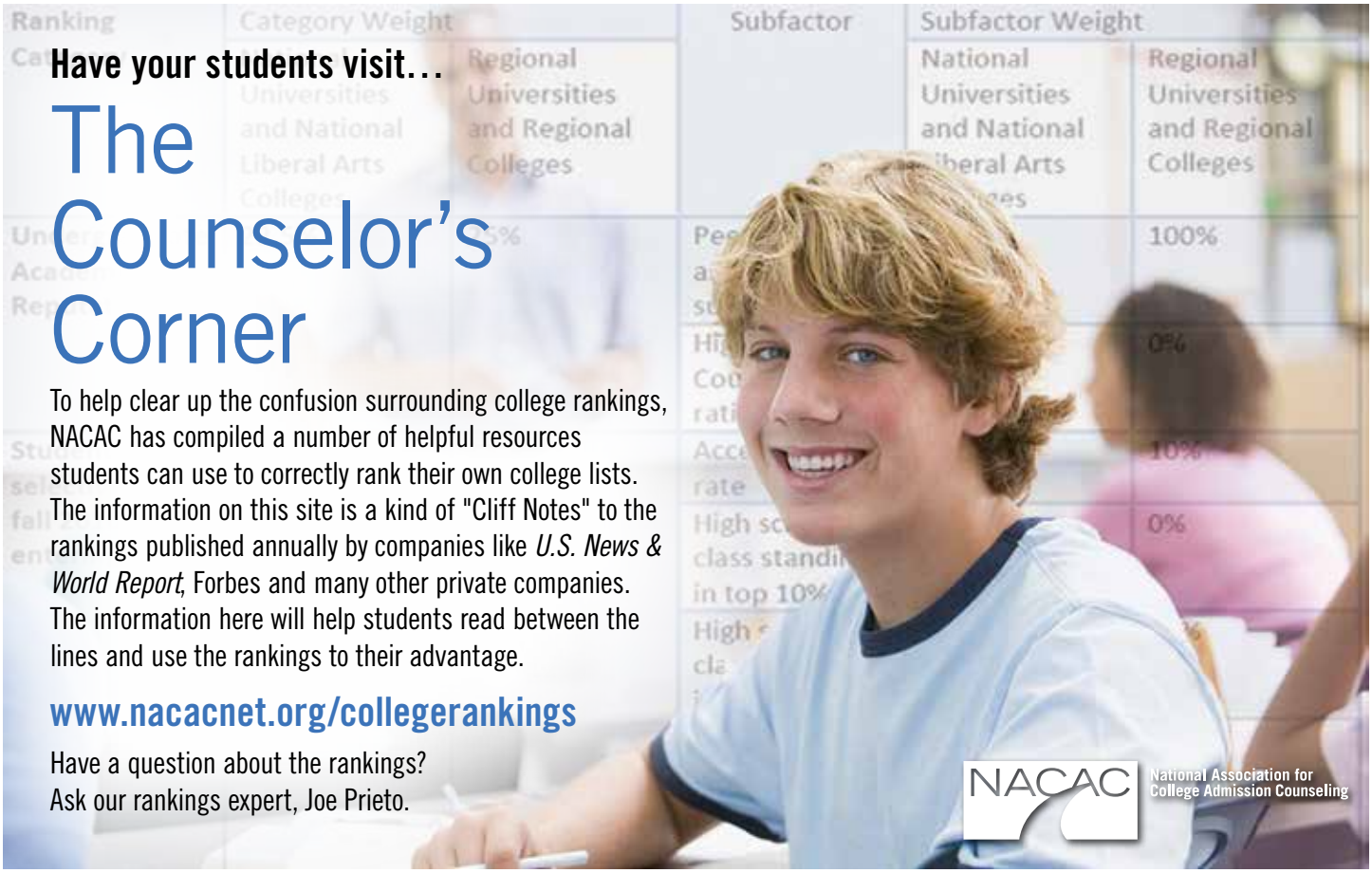
this finding is that very few students have a dozen or more scores through which admission officers must sift during the admission process. In that same vein, 35 percent of multiple-SAT-takers in this sample attain their highest individual section scores on the same administration, meaning that for this sub-group of examinees, the highest individual section and highest single administration score sets are identical. Despite these findings and because many students do take the SAT more than once, a score set policy is necessary for the sake of fairness.

The main goal of this study was to determine if the choice of score set policy would lead to differences in the predictive validity of SAT scores for FYGPA. The results generally point to the same conclusion: alone or when used in conjunction with HSGPA, the various sets of SAT scores used in admission processes have nearly the same linear relationship with students' FYGPAs. This is not surprising given that the differences in mean SAT scores under different score set policies are relatively small. This is especially true of the two most common score set policies: highest individual sections and highest single administration, where difference in means for those two score sets across all institution subgroups and all student subgroups—except number of testing occasions—is less than 20 points. The implications of this finding are that no matter which score(s) an institution chooses to require for admission, the predictive validity of the SAT is not undermined. In other words, there is little adverse impact in terms of predictive validity associated with an institution's admission officers using one score set, as opposed to another. The highest individual sections score set has about the same predictive validity as any other score set for the groups considered in this study, validating what appears to be common practice at many institutions. So which set of scores should be used in the admission process? Each institution's admission officers would do well to consider these results, as well as historical data and other evidence specific to their applicant pool. Rigorous empirical research in conjunction with thoughtful conversation about the institution's mission and goals can help inform which set of scores will best serve the institution.

ENDNOTES

[1] For more information on institution policies for the use of SAT scores in the admission process, see College Board (2010b).

[2] Based on institution-supplied score use practices (referred to as "score set policies" in this study), 66 percent of four-year institutions for which data were available reported that they use "Highest Section" (i.e., highest individual sections) and 15 percent reported using "Highest Sitting" (i.e., highest single administration) with the remainder (19 percent) reporting that they request all SAT scores (for more detail, see College Board, 2010a).



BRIAN F. PATTERSON is an assistant research scientist at The College Board. He holds an M.A. in Statistics and is completing his Ph.D. dissertation in Measurement and Evaluation, both from Columbia University (NY). His applied work focuses on the validity of educational tests and programs, and his research interests include multilevel modeling and propensity score methods.

DR. PETER SWERDZEWSKI is a fellow at the Regents Research Fund specializing in educational assessment. He holds a M.Ed. in Counseling Psychology and a Ph.D. in Assessment and Measurement, both from James Madison University (VA). His research interests include structural equation modeling, assessment and public policy, and the assessment of non-cognitive constructs.

DR. KRISTA D. MATTERN is an associate research scientist at The College Board and is primarily responsible for conducting research on test validity. Krista received her Ph.D. in Industrial/Organizational Psychology with a minor in Quantitative Psychology from the University of Illinois, Urbana-Champaign in 2006. She joined The College Board in June 2006.

REFERENCES

AERA, APA, and NCME. (1999). *Standards for educational and psychological testing.* Washington, DC: Author. American Educational Research Association/American Psychological Association/National Council on Measurement in Education.

Agronow, S. and Studley, R. (2007, November). *Prediction of college GPA from new SAT test scores – A first look.* Paper presented at the Annual Meeting of the California Association for Institutional Research (CAIR), Monterey, CA.

Alderman, D. L. (1981). Student self-selection and test repetition (College Board Research Note No. 81-5). New York: The College Board. Retrieved from: http://research.collegeboard.org/rr1981-5.pdf.

Camara, W. J. and Echternacht, G. (2000). The SAT I and high school grades: Utility in predicting success in college (Office of Research and Development Research Notes No. RN-10). New York: The College Board. Retrieved from: http://research.collegeboard.org/rn2000-10.pdf.

Clinedinst, M. E and Hawkins, D. A. (2010). *NACAC State of College Admission 2010.* Arlington, VA: National Association for College Admission Counseling.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, N.J: L. Erlbaum Associates.

Cornwell, C. M., Mustard, D. B., and Van Parys, J. (2008). *How does the new SAT predict academic achievement in college?* Manuscript submitted for publication. Retrieved on March 17, 2009, from http://www.terry.uga.edu/~cornwl/research/newsat.pdf.

Kobrin, J. L., Patterson, B. F., Shaw, E. J., Mattern, K. D., and Barbuti, S. M. (2008). Validity of the SAT® for Predicting First-Year College Grade Point Average (College Board Research Rep. No. 2008-5). New York, NY: The College Board. Retrieved from: http://research.collegeboard.org/rr2008-5.pdf.

Lawley, D. N. (1943). A note on Karl Pearson's selection formulae. *Proceedings of the Royal Society of Edinburgh*, 62(1), 28-30.

Mattern, K. D., Patterson, B. F., Shaw, E. J., Kobrin, J. L., and Barbuti, S. M. (2008). Differential Validity and Prediction of the SAT® (College Board Research Rep. No. 2008-4). New York, NY: The College Board. Retrieved from: http://research.collegeboard.org/rr2008-4.pdf.

NACAC. (2008). *Statement of principles of good practice: Approved by the 2008 Assembly.* Arlington, VA: National Association for College Admission Counseling.

NACAC. (2010). *Statement of principles of good practice: Approved by the 2010 Assembly.* Arlington, VA: National Association for College Admission Counseling.

Patterson, B. F., Mattern, K. D., and Kobrin, J. L. (2009). Validity of the SAT for Predicting FYGPA: 2007 SAT Validity Sample (College Board Statistical Rep. No. 2009-1). New York, NY: The College Board. Retrieved from: http://research.collegeboard.org/sr2009-1.pdf.

Patterson, B. F. and Mattern, K. D. (2011). Validity of the SAT for Predicting FYGPA: 2008 SAT Validity Sample (College Board Statistical Rep. No. 2011-5). New York, NY: The College Board. Retrieved from: http://research.collegeboard.org/sr2011-5.pdf.

Patterson, B. F. and Mattern, K. D. (2012). Validity of the SAT for Predicting FYGPA: 2009 SAT Validity Sample. New York, NY: The College Board. Retrieved from: http://research.collegeboard.org/sr2012-3.pdf.

Powers, D. E. (2004). Validity of graduate record examinations (GRE) general test scores for admission to colleges of veterinary medicine. *Journal of Applied Psychology,* 89(2), 208-219.

The College Board. (2006). 2006 College-bound seniors: *Total group profile report.* New York: The College Board. Retrieved from: http://media.collegeboard.com/digitalServices/pdf/research/cb-seniors-2006-national-report.pdf.

The College Board. (2009). *The College Board college handbook 2009 (46th ed.).* New York: The College Board.

The College Board. (2010a). *SAT® score-use practices by participating institution.* Retrieved, from http://professionals.collegeboard.com/profdownload/sat-score-use-practices-list.pdf.

The College Board. (2010b). *Guidelines on the uses of College Board test scores and related data.* New York: The College Board. Retrieved from: http://professionals.collegeboard.com/profdownload/guidelines-on-uses-of-college-board-test-scores-and-data.pdf.

Whetstone, C. (2008, October 10). Averaging SAT scores. Message posted to NACAC E-List, achieved at http://peach.ease.lsoft.com/archives/nacac.html.

For the following appendices, contact the authors at bpatterson@collegeboard.org.

Appendix A. Raw Correlations[a] of FYGPA with SAT Score Sets and HSGPA by Institutional Characteristics and Number of Times Students Took the SAT

Appendix B. Corrected (Raw) Correlations[a] of FYGPA with HSGPA and the First SAT Taken across Various Institutional Characteristics and Number of Times Students Took the SAT

Appendix C. Corrected (Raw) Correlations[a] of FYGPA with HSGPA and the Latest SAT Taken across Various Institutional Characteristics and Number of Times Students Took the SAT

Appendix D. Corrected (Raw) Correlations[a] of FYGPA with HSGPA and the Average SAT Taken across Various Institutional Characteristics and Number of Times Students Took the SAT

Appendix E. Corrected (Raw) Correlations[a] of FYGPA with HSGPA and the Highest Single SAT Administration by Institutional Characteristics and Number of Times Students Took the SAT

Appendix F. Corrected (Raw) Correlations[a] of FYGPA with HSGPA and the Highest Individual SAT Sections by Institutional Characteristics and Number of Times Students Took the SAT