



National Association for
College Admission Counseling
Guiding the way to higher education

FOUNDATIONS OF STANDARDIZED ADMISSION TESTING

Richard J. Noeth, Ph.D.

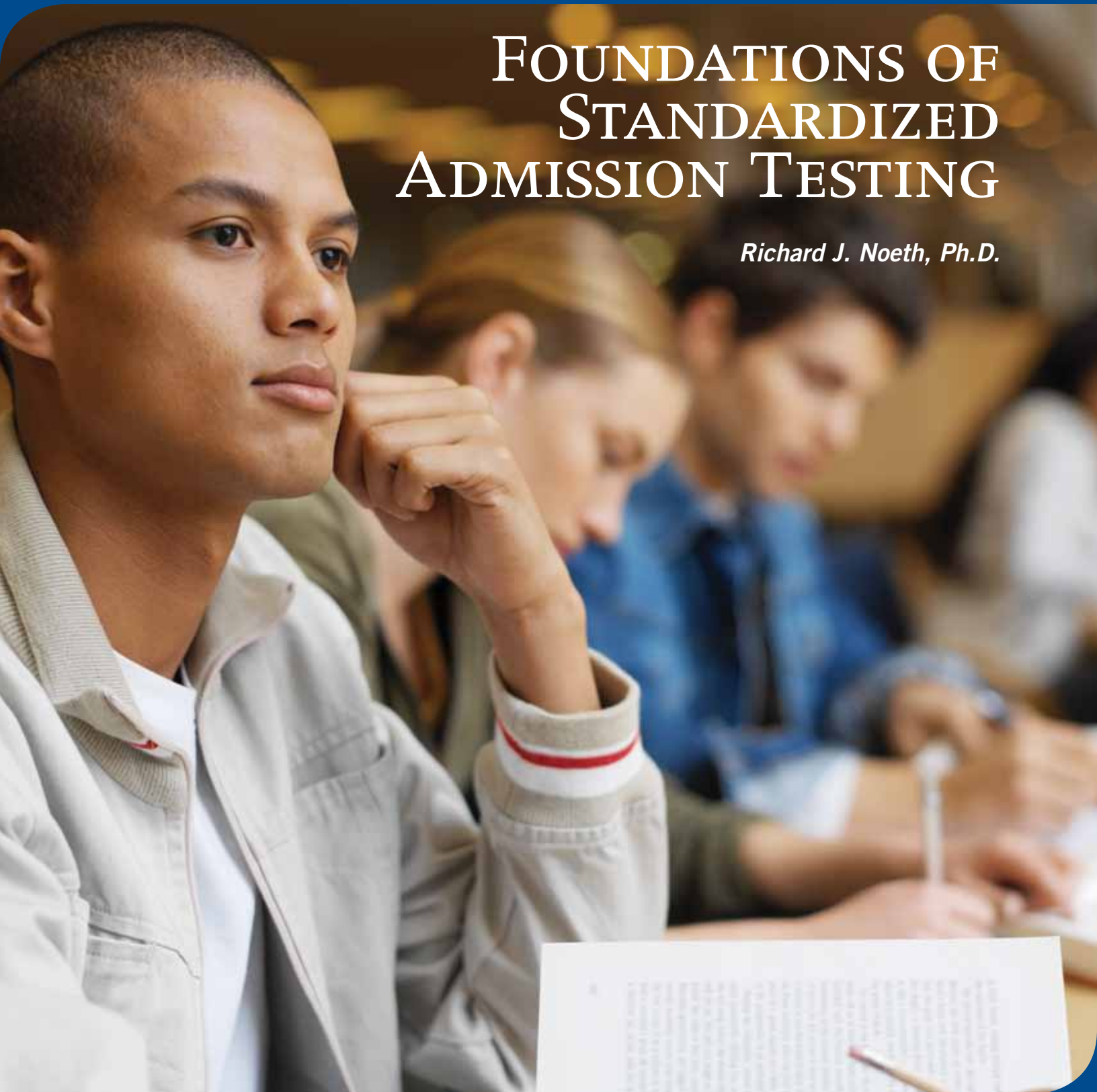


TABLE OF CONTENTS

ACKNOWLEDGEMENTS	I
----------------------------	---

I. INTRODUCTION

National Research Council Report on Admission Testing.	5
Commission Recommendations	5
Foundations of Admission Testing	6
Learning Assignments.	6

II. ETHICAL AND PROFESSIONAL STANDARDS IN ADMISSION TESTING

Standards for Educational and Psychological Testing	9
Code of Fair Testing Practices in Education.	10
NACAC Statement of Principles of Good Practice	11
Test-Taker Rights Regarding Admission Testing	11
Learning Assignments.	11

III. OBJECTIVES OF ADMISSION TESTING

Admission Testing from the Beginning	15
Admission Models and Standardized Testing	16
Admission Test Information – Benefits and Limitations.	18
Learning Assignments.	19

IV. THE CONTENT AND PROPERTIES OF ADMISSION TESTS

Admission Test Content.	23
Reliability of Admission Tests	27
Admission Test Difficulty.	28
Admission Test Scoring.	29
Learning Assignments.	30

V. INTERPRETIVE CONCEPTS AND PRINCIPLES

Interpreting Group Admission Test Results	33
Interpreting Admission Test Results for Individual Students	35
Concordance between Admission Tests	37
Learning Assignments.	37

VI. USING TESTS TO PREDICT COLLEGE SUCCESS

Defining and Describing Predictive Validity	41
The Predictive Validity of Admission Tests.	42
Differential Validity and Differential Prediction	43
Institutional Test Validity Evidence.	45
Learning Assignments.	45

VII. STUDENT PREPARATION FOR ADMISSION TESTS

Test Preparation and Test Coaching Programs	49
Effects of Test Preparation and Test Coaching	50
Implication of Test Score Increases	50
Learning Assignments.	51

VIII. GROUP DISPARITIES IN ADMISSION TEST SCORES

Developing Fair Admission Tests	55
Differential Test Performance	56
Examining Admission Test Group Differences	56
Interpreting Group Difference Information.	58
Learning Assignments.	58

ADDITIONAL ADMISSION TEST INFORMATION RESOURCES . . 61

ADMISSION TESTING SELF-ASSESSMENT MODULE

Questions	63
Answers.	79

I

INTRODUCTION

Mission: The National Association for College Admission Counseling (NACAC) is dedicated to supporting the work of counseling and enrollment professionals as they advance student access to postsecondary education. NACAC is committed to promoting high professional standards that foster ethical and social responsibility.

Given this Mission and the highly-visible role that standardized tests play in undergraduate admission (including on-going concerns about a range of testing issues), NACAC appointed a Commission on the Use of Standardized Tests in Undergraduate Admission¹ to make recommendations about admission test use. The commission sought to determine how NACAC, institutions of higher education and other stakeholders interpret and implement policies or practices with regard to the use of standardized tests in undergraduate admission.²

National Research Council Report on Admission Testing

The background for the work of the commission was based on the core observations of the National Research Council (NRC) in its 1999 report, *Myths and Tradeoffs: The Role of Tests in Undergraduate Admissions*.³ The extent to which current practice reflects the conclusions of the NRC report provided the framework for commission discussions. This report stated the clear benefits of tests were that:

can lead admission officers to consider acceptance for a student who would otherwise be rejected.

The NRC report weighed these benefits with the concern that test scores were also used in ways that were not in line with their design or stated purposes, were beyond their technical capacities or were detrimental to important widely-shared goals of a fair, open and effective admission process. These “tradeoffs” thus led the NRC to offer four recommendations to postsecondary institutions:

1. Admission policies and practices should be derived from and clearly linked to an institution’s overarching intellectual and other goals.
2. The use of test scores in the admission process should serve those institutional goals.
3. Admission policies, and their relationship to institutional goals, should be clearly articulated for the public so that students can make informed decisions about whether to apply.
4. Postsecondary institutions should review their use of test scores in the admission process and, if necessary, take steps to eliminate misuses of scores; and specifically avoid treating scores as more precise and accurate measures than they are and not rely on them for fine distinctions among applicants.

The commission emphasized that a “one-size-fits-all” approach for the use of standardized tests in undergraduate admission does not reflect the realities facing the nation’s many and varied colleges and universities. The commission focused on five critical issues that affect all institutions and offered a series of recommendations in its final report.

1. Given the variety that characterizes the US educational system and the variability of curricula, grading standards and course content, standardized tests are an efficient source of comparative information for which there is currently no substitute.
2. Standardized tests can be provided at a relatively low cost to students and offer valuable efficiencies to institutions that must review thousands of applications.
3. Standardized tests provide students with an opportunity to demonstrate talent; and for those whose academic records are not particularly strong, a high score

Commission Recommendations

The commission emphasized that a “one-size-fits-all” approach for the use of standardized tests in undergraduate admission does not reflect the realities facing the nation’s many and varied colleges and universities. The commission focused on five critical issues that affect all institutions and offered a series of recommendations in its final report. Among them were that NACAC:

- Establish opportunities for colleges and high schools, as well as college admission counseling professionals,

¹ Hereafter referred to as the commission.

² National Association for College Admission Counseling. (2008). *Report of the Commission on the Use of Standardized Tests in Undergraduate Admission*. Arlington, VA: Author. (http://www.nacacnet.org/PublicationsResources/Research/Documents/TestingCommission_FinalReport.pdf)

³ National Research Council. (1999). *Myths and Tradeoffs: The Role of Tests in Undergraduate Admissions*. Washington, DC: National Academies Press.

II

ETHICAL AND PROFESSIONAL STANDARDS IN ADMISSION TESTING

The Statement of Principles of Good Practice (SPGP) identifies ethical college admission as the cornerstone of NACAC. A principal goal is to protect students' rights in the transition to postsecondary education through monitoring and enforcing ethical standards and practices. The commission has reinforced this goal with regard to admission tests by issuing the following recommendation for SPGP amendment:

Familiarity with Standards for Educational and Psychological Testing. College admission officials in charge of setting admission test policy should be familiar with and adhere to the *Standards for Ethical and Psychological Testing*, particularly with respect to test score use and interpretation, test bias and score differences between subgroups.

This section focuses on three publications related to ethical and professional standards in admission testing: (1) the *Standards for Educational and Psychological Testing*,⁸ (2) the *Code of Fair Testing Practices in Education* and (3) the NACAC Statement of Principles of Good Practice. This is followed by a brief discussion of test-taker rights throughout the testing and score reporting process.

Standards for Educational and Psychological Testing

The *Standards for Educational and Psychological Testing** is a comprehensive, 190-page book (including a glossary) written



of the text is also directed toward ethical standards for those who develop tests.

The *Standards* are separated into three parts and the most relevant chapters for admission professionals are:

College admission officials in charge of setting admission test policy should be familiar with and adhere to the *Standards for Ethical and Psychological Testing*, particularly with respect to test score use and interpretation, test bias and score differences between subgroups.

at an easy-to-understand, non-technical level. Its goal is to promote the sound and ethical use of tests and to provide a basis for evaluating the quality of testing practices.

The *Standards* is based on the premise that effective testing requires that all participants possess the knowledge, skills and abilities relevant to their roles in the testing process, as well as awareness of personal and contextual factors that may influence the process. However, because the *Standards* cover all types of assessment and testing (e.g., educational testing, psychological testing, employment testing), not all chapters are appropriate for admission test users. A significant portion

- **Part I: Test Construction, Evaluation and Documentation.** Chapters 1 (Validity), 2 (Reliability and Errors of Measurement), 4 (Scales, Norms and Score Comparability), and 5 (Test Administration, Scoring and Reporting).
- **Part II: Fairness in Testing.** Chapters 7 (Fairness in Testing and Test Use), 8 (The Rights and Responsibilities of Test Takers), 9 (Testing Individuals of Diverse Linguistic Backgrounds), and 10 (Testing Individuals with Disabilities).
- **Part III: Testing Applications.** Chapters 11 (The Responsibilities of Test Users) and 13 (Educational Testing and Assessment).

⁸ American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for Educational and Psychological Testing*. Washington, DC: American Educational Research Association. *The *Standards for Educational and Psychological Testing* is published by the American Educational Research Association (AERA) and can be ordered directly from AERA at www.aera.net for a non-member price of \$49.95 plus shipping.

III

OBJECTIVES OF ADMISSION TESTING

College admission tests have been part of the educational landscape for over one hundred years. Yet issues continue to be raised, often cyclically, about their use, appropriateness and fairness. This section provides a background on test use in undergraduate admission. It describes their growing use and how tests are positioned within several undergraduate admission models. It also explores test benefits and limitations (including uses and misuses) and institutional responsibilities regarding formal admission policies and practices related to testing.

Admission Testing from the Beginning

Examining students to determine their readiness to successfully pursue postsecondary education began in the US at the beginning of the 20th century. The College Board was created in 1900 to respond to the disarray of institutional admission examinations. It developed a standard set of essay tests for use across multiple colleges and universities that were first administered in 1901.

The foundation for the standardized admission testing programs that are used today began about 25 years later with the predominantly multiple-choice College Board Scholastic Aptitude Test (SAT). The ACT began in 1959 as a competitor to the SAT. Although both exams began with geographical strongholds, they are now virtually accepted and used interchangeably for admission throughout the United States.

Commensurately, the role that admission tests play within the transition from secondary to postsecondary education has also grown. The *State of College Admission*¹⁴ reports that from 1993 through 2008, the percentage of colleges attributing “considerable importance” to test scores within admission decision factors generally increased (Table 1).

Table 1. Percentage of Colleges Attributing Considerable Importance to Test Scores in the Admission Decision

Year	Percent	Year	Percent
1993	46	2001	52
1994	43	2002	57
1995	47	2003	61
1996	48	2004	60
1997	50	2005	59
1998	51	2006	60
1999	54	2007	59
2000	58	2008	54

Besides the growth in the importance of test scores in admission decisions, it is equally important to view their relative

College admission tests have been part of the educational landscape for over one hundred years. Yet issues continue to be raised, often cyclically, about their use, appropriateness and fairness.

The volume of students who take admission tests has grown dramatically since the early days of these exams. These increases have often been associated with legislative, social and/or economic benchmarks in American society. Most recently, for the 2008 college-bound class, 1,518,859 (in 1998, SAT I volume was 1,172,779)¹³ completed the SAT Reasoning Test; the number of ACT test takers was 1,421,941 (it was 995,039 10 years earlier).

position within the scope of other admission-consideration factors. These other factors have traditionally spanned the range of academic experiences (e.g., grades, strength of curriculum), essays, recommendations, and extracurricular activities. Table 2 shows the top 10 factors of “considerable” or “moderate” importance for admission decisions in 2008. Consistent with its upward trend in importance, testing represents a significant and sizeable factor in admission decision making.

¹³ The sources for these test-taker volumes were the SAT and ACT annual and archived college-bound senior and national score data (respectively) found at <http://professionals.collegeboard.com/data-reports-research/sat/archived> (for the SAT) and <http://www.act.org/news/data.html> (for the ACT).

¹⁴ Hawkins, D.A., & Clinedinst, M.E. (2009). *State of College Admission*. Arlington, VA: National Association for College Admission Counseling. (<http://www.nacacnet.org/PublicationsResources/Research/Documents/09StateofAdmission.pdf>)

IV

THE CONTENT AND PROPERTIES OF ADMISSION TESTS

The *Standards* and the *Code* strongly encourage those who select tests as part of a college admission policy, who interpret test results and/or who utilize test information in the decision-making process to be knowledgeable about what the tests actually measure. This stems from concerns about test fairness, appropriate test use and professional responsibilities to test takers and to institutions that rely on the informed judgment of their test users.

This section will focus on the specific content of admission tests. It will describe approaches to test content; then focus on how the SAT and ACT* function in terms of the knowledge and skills they measure, and how this knowledge and these skills are assessed



validly predict postsecondary performance, have remained remarkably similar.

SAT Reasoning Test. The SAT is designed to test basic knowledge of subjects students have learned in the classroom—reading, writing and mathematics—in addition to how they think, solve problems and communicate. The SAT is not aligned with a specific high school curriculum *per se*. It examines how well students apply the skills and knowledge they have attained and developed both inside and outside of the classroom over their educational careers; all with the function of assessing critical thinking skills needed to meet postsecondary academic challenges.

The SAT accomplishes these tasks through the assessment of critical thinking and problem-solving skills in three content areas:

- Critical Reading
- Mathematics
- Writing

These content areas comprise the present three hour and 45 minute SAT Reasoning Test, which consists of 10 separately-timed sections. (One section is a 25-minute unscored component that is used to try out new test questions and to ensure that scores on new editions of the SAT are comparable to scores on earlier editions.)

The original and ongoing philosophies in terms of the developmental purposes and actual contents of the SAT and ACT have remained distinctive. However, the statistical relationships between these two tests, as well as their ability to validly predict postsecondary performance, have remained remarkably similar.

Admission Test Content

A principal use of test information in admission has been to help colleges and universities determine which students are prepared to successfully meet the institution's academic challenges. The original and ongoing philosophies in terms of the developmental purposes and actual contents of the SAT and ACT have remained distinctive. However, the statistical relationships between these two tests, as well as their ability to

SAT Critical Reading. There are three Critical Reading sections that require a total of 70 minutes of testing time (two of 25 minutes and one of 20 minutes). The number of questions by type and content category type is shown in Tables 3 and 4. Critical Reading questions are all multiple choice and they assess reading skills that include:

- Identifying main and supporting ideas.
- Determining the meaning of words in context.
- Understanding authors' purposes.
- Understanding the structure and function of sentences.

*Information on the SAT Reasoning Test and on the ACT has been provided by the College Board and ACT directly for this textbook. The 2008-09 testing year publications consist of the *SAT Program Handbook* and the *ACT User Handbook*. Also utilized are *The ACT Technical Manual*, *The ACT Writing Test: Preliminary Technical Report*, selected downloads from the respective Web sites (collegeboard.com and act.org), and technical information provided directly by College Board and ACT staff.

V

INTERPRETIVE CONCEPTS AND PRINCIPLES

There are a number of measurement concepts and principles that test users must understand to adequately and accurately interpret and apply admission test information. It is important that admission professionals understand the conceptual and statistical language of testing to correctly and responsibly utilize test information in the decision-making process as well as to communicate it effectively to their constituencies. *The Statement on Counselor Competencies* (www.nacacnet.org/AboutNACAC/Policies/Documents/CounselorCompetencies.pdf) states:

Secondary school counselors and college admission counselors should be able to interpret test scores and test-related data to students, parents, educators, institutions, agencies, and the public.

This section will focus on a basic set of fundamental interpretive and statistical concepts and principles for admission testing. These will include the interpretation of admission test information for groups as well as for individual test takers. It will specifically describe the interpretation of SAT and ACT scores and include a brief presentation of their concordance with each other.

Interpreting Group Admission Test Results

Admission professionals receive and review an array of test in-

Measures of Central Tendency. Measures of central tendency (measures of the center of a distribution) are meant to reflect the most accurate and understandable methods used to describe the test performance of a group of individuals. There are three measures of central tendency that can be used to describe a group of individuals: the mean, median and mode.

The mean is the most frequently used of the three measures; likely because it is most easily understood, calculated and conveyed. The mean is simply the arithmetic average of a distribution of test scores. It is calculated by adding all of the scores for a student cohort (e.g., applicants for a particular program) and dividing this total number by the number of students in the cohort.

Admission professionals and the public receive information about the average college-bound senior test scores on the SAT or ACT each year. These are the mean scores for the respective cohorts and are calculated by adding all of the SAT section or ACT section (or Composite) scores and then dividing these totals by the number of SAT or ACT test takers. Table 15 presents a brief hypothetical distribution of test scores for two groups. The mean (average score) for the SAT Mathematics test is 531 and the mean for the ACT Mathematics test is 22.1.

The median is the next used measure of central tendency. The median is the middle-most test score for a given group when all scores have been arranged in order of size. The me-

Admission professionals receive and review an array of test information about students. They make judgments and analyses using this information about the preparedness levels of these students.

formation about students. They make judgments and analyses using this information about the preparedness levels of these students. They often convey these judgments and analyses about academic readiness to other interested parties—including educators, administrators and the public.

There are generally two essential ways to analyze and convey test information about student groups. One involves measures of central tendency used to describe, organize and summarize the full distribution (the complete set or complete group) of test scores under consideration. The other involves measures of dispersion used to provide information about the spread of scores for a given cohort or set of cohorts.

dian is the exact point that cuts in half the entire set of scores—the score at which half of the individuals fall above and at which half of the individuals fall below. If there is an even number of scores in the distribution, the median is the arithmetic average of the two scores in the middle of the distribution. Unlike the mean, the median is less affected by extreme scores (i.e., outliers).

Looking at the hypothetical distributions of scores in Table 15, the median score for the SAT Mathematics test is 510 and the median for the ACT Mathematics test is 21. Each of these scores is exactly the middle score of the hypothetical set of SAT Mathematics and ACT Mathematics scores (each bisects the respective distribution of scores).

VI

USING TESTS TO PREDICT COLLEGE SUCCESS

The first chapter of the *Standards* focuses on test validity and characterizes it as the most fundamental consideration in developing and evaluating tests. Validity refers to the degree in which evidence and theory support the interpretations of test scores involved in the proposed uses of tests. In other words, what scientific evidence exists that supports the different ways in which tests are used?

One crucial type of validity for tests used in admission decision making is predictive validity. This section focuses on the predictive validity function of admission tests. It describes the meaning, process and interpretation of predictive validity and includes a discussion of the statistical concepts involved. It covers other factors that comprise the prediction of postsecondary success, the issue of test validity for different student population groups and the need for institutions to establish their own validity evidence.

Defining and Describing Predictive Validity

One of the hallmarks of tests used in undergraduate admission is their role in predicting success in college. This is a principal foundation of test use for postsecondary admission in that tests can help predict who will be successful and who will likely struggle with an institution's academic challenges. This information, along with an array of other factors, helps test users make informed judgments about applicants.

Predictive validity is essentially the relationship between a predictor and a criterion. It is not causal in that the predictor does not cause the criterion to occur at a later date. It is correlational in that the predictor is related to (associated with) the criterion. In the case of college admission, predictive validity might typically involve a test score (SAT-ACT) as the predictor and First-Year College Grade Point Average (FGPA) as the criterion. This basic one-to-one function is typically expanded to include other predictors that might include a student's High School Grade Point Average (HSGPA) and perhaps class rank. The criterion, as well, might be expanded to include college persistence (e.g., into the second year) and college graduation.

Predictive validity is expressed as a statistical relationship between the predictor(s) and criterion. This relationship is expressed as a correlation coefficient that ranges from 0.0 (no relationship) to 1.00 (perfect relationship). Regarding admission tests, if there were no relationship between performance on the test (the predictor) and FGPA (the criterion), the correlation would be 0.0. If the test perfectly predicted FGPA, the predictive validity coefficient would be 1.0.

Unlike reliability coefficients (discussed in Section 4), which should be in the .80–.90 range for high stakes admission tests, expectations for satisfactory validity coefficients tend to be lower—typically in the .30–.50+ range. (Correlations of .10,

.30 and .50 or higher typically demonstrate small, medium and large predictor-criterion relationships, respectively.) There are multiple reasons for lower validity coefficient expectations:

- **Predictive Factors.** There are a broad range of predictive factors that can relate to criterion measures. For example, academic, social, economic, and personal issues might all, to some extent, potentially impact a student's academic success during the first year of college. (There are likely even more factors over time that might influence a student's persistence and graduation.)
- **Criterion Variability.** There is considerable variability in postsecondary success criteria. For example, FGPA will likely be a function of courses taken, course difficulties, grading standards, instructional demands, and similar factors within and across different institutions. The ability of factors such as admission test scores and HSGPA to explain FGPA—when these criterion influences (e.g., things affecting FGPA) will likely vary within and across institutions—is expectedly somewhat limited.
- **Restriction of Range.** The predictive validity coefficient for a given college will be a restricted value due to the fact that the validity of the predictors (e.g., test score, HSGPA) related to the criterion (e.g., FGPA) is calculated on only those students who were admitted and chose to attend that institution. The validity calculation does not include those who did not gain admission or who were admitted but did not attend. Thus the range of potential test scores and HSGPA (the predictors) and the range of the potential FGPA (the criterion) are restricted or limited.

The predictive validity coefficient is often interpreted by a related statistic. Specifically, it is squared to provide the proportion or percent of the variability in the criterion (e.g., FGPA) that is explained by, related to, or associated with the predictor(s) (e.g., admission test score and/or HSGPA). For example, if the predictive validity correlation were .50 for a test score and HSGPA combined (the predictors) as they relate to FGPA (the criterion), then the proportion of the criterion (FGPA) explained by or associated with the predictors (test score, HSGPA) would be 25 percent (the .50 predictive validity coefficient squared to equal 25 percent).

If 25 percent of the postsecondary success criterion (FGPA) is explained by, related to, or associated with the combined predictors of test score and HSGPA, then 75 percent of the criterion (FGPA) must be explained by, related to, or associated with other (non-test score and non-HSGPA) factors. In the case of FGPA, these other factors might include social, interpersonal,

VII

STUDENT PREPARATION FOR ADMISSION TESTS

Issues of test preparation and test coaching are often at the forefront of discussions about admission test fairness. The concern is whether students with substantial family incomes and students with access to schools with resources to offer comprehensive test preparation activities gain an advantage over those students who could not afford to engage in such programs or whose schools lacked the resources to offer them. The premise is that test preparation and test coaching programs can lead to significant score increases on admission tests and thus advantage students throughout the admission process.

This section will discuss the range of test preparation activities and programs that currently exist; including their focus, intensity and required level of commitment. It will describe the documented effects of test preparation and test coaching in terms of score increases on admission tests. It will then analyze the potential ramifications of possible test score increases due to test preparation activities, particularly in terms of admission decision making.

Test Preparation and Test Coaching Programs

There is a broad array of programs and activities that supports student preparation for college admission tests. These can include, among others, those that are: (1) commercial (and well publicized), (2) based in secondary or postsecondary institu-

- Familiarity with the test (e.g., becoming familiar with test instructions, question types and question format, test section time limits, and using a calculator to solve questions)
- Reviewing content relevant to a test (e.g., reviewing the content of the mathematical or science portion of a test)
- Learning testwiseness skills (e.g., strategies for guessing, eliminating choices in multiple-choice questions and using time efficiently)

The review refers to the distinction between short-term programs that emphasize drill and testwiseness, and longer-term programs that resemble supplementary academic instruction. It also distinguishes between commercial test coaching programs that tend to emphasize score increases and shorter-term school-based programs that focus more on familiarity with the test. Similarly, a NACAC Discussion Paper²⁷ describes three elements that are generally at the core of test preparation programs:

- A review of test content
- Practice on test questions
- Orientation to the format of the test (i.e., development of “testwiseness”)

Both the College Board and ACT affirm that preparation for their tests is important for all students. They acknowledge

The concern is whether students with substantial family incomes and students with access to schools with resources to offer comprehensive test preparation activities gain an advantage over those students who could not afford to engage in such programs or whose schools lacked the resources to offer them.

tions, (3) situated in community organizations, (4) provided by one-on-one tutors, or (5) online. This can also include independent preparation by students working on their own. The materials utilized can be commercially published, provided free or for a fee by test publishers (the College Board or ACT), locally-developed by teachers for use in the classroom, and/or be interactive and electronically delivered.

This array of test preparation activities is also embodied in the different methods, objectives, focus, and duration of the programs available. One review²⁶ describes these elements as focusing on:

that such preparation will (minimally) improve test scores, as well as help students to perform their best by becoming familiar with test content, format, timing, directions, etc. They encourage students to prepare for their tests and provide a range of free and fee-based paper, electronic and online materials to support this preparation. These support materials include an overview of the tests, test familiarization information, practice questions, test-taking strategies, and explanations and guides with reasons for correct and incorrect choices. Students independently engaged in these activities represent the more informal, student-driven and self-guided mode of test preparation.

²⁶ Powers, D.E., & Camara, W.J. (1999). *Coaching and the SAT I*. (College Board RN-06). New York: The College Board.

²⁷ Briggs, D.C. (2009). *Preparation for College Admission Exams*. (A Discussion Paper for the National Association for College Admission Counseling.) Arlington, VA: National Association for College Admission Counseling.

VIII

GROUP DISPARITIES IN ADMISSION TEST SCORES

The issue of fairness to all applicants is a bedrock of the admission process. Given the major role of standardized testing in this process, the empirically-based reassurance, acceptance and belief that admission tests are fair to all test takers cannot be overestimated.

Test score differences have existed among racial/ethnic, gender, best language, and socioeconomic groups for a long time. These differences continue to raise ongoing concerns about the possibility of bias in admission tests. This section will focus on three aspects of perceived test fairness. One will be the development of admission tests and how this process is undertaken to ensure fair and unbiased tests. The second will examine group data that display the actual differences in test scores. The third will consider how these disparities should be examined and interpreted within the concept of fairness in admission decision making.

Developing Fair Admission Tests

The *Standards* and *Code* focus directly on issues of fairness and bias—including the development of the test. The *Standards* state that bias arises when deficiencies in a test result in different meanings for scores earned by members of different groups. The *Code* raises the issue that there needs to be evidence to ensure that differences in performance are related to the skills being assessed.

To avoid the possibility of question bias, two notable steps in the test development process have been in place for a number of years. One is the sensitivity or fairness review which ensures that the content of a given question is not disturbing or offensive to certain groups of students in some particular way, which might inadvertently affect their performance on that question.

This sensitivity review reflects the guideline included in the *Code* which calls for the avoidance of potentially offensive content or language when developing test questions. This review is typically performed at several stages in the test development process including both early in question development and then closer to the question being finalized, especially if it has been modified during the process. Sensitivity reviews are generally performed by committees of high school and college faculty who come from geographically and ethnically diverse backgrounds.

The second safeguard against question bias in the test development process is the application of the Differential Item Functioning (DIF) screening. DIF is a detailed statistical procedure that allows test developers to determine whether equally skilled members of different groups (e.g., females-males, Hispanic students-Caucasian students) have statistically different rates of correct responses on particular questions. The purpose is to identify questions that may be inappropriately problematic for a student group because they may contain material that is

Test score differences have existed among racial/ethnic, gender, best language, and socioeconomic groups for a long time. These differences continue to raise ongoing concerns about the possibility of bias in admission tests.

One area of concern is the actual content of the test. The issue is whether the test questions (also called test items), as encountered by different members of the test-taking population, are biased. Specifically, is each test question fair to all test takers? Or does its content and presentation negatively affect the performance of different groups in any other way than to correctly assess the academic skill being measured by that question? Is there something about the wording of a question, its language and/or its use of a particular topic that affects various population groups differently in terms of their performance on that question; and is this effect unrelated to what is actually being measured by the question?

irrelevant to the construct being assessed and/or may be more familiar to some groups than to others.

The DIF procedure is carried out for all test questions, and questions are flagged (indicating the need for further scrutiny) if they contain a certain DIF level. Once flagged, questions are reviewed by panelists from a variety of backgrounds who are experts in the particular subject area to determine, if possible, the likely cause for the DIF. If the question is considered acceptable and valid despite its differential functioning, it remains on the test. If the question is determined to be biased, it is modified or eliminated.

ADDITIONAL ADMISSION TEST INFORMATION RESOURCES

The following five statistical and educational testing and measurement textbooks are among those currently used in graduate classrooms around the country in such courses as basic statistical methods and introductory testing and measurement. They contain and discuss all of the statistical concepts and applications covered in this textbook. In addition to the publications and materials listed in the **Learning Assignments** under each textbook section, these additional publications can serve as further resources to textbook users and their colleagues.

- *Statistics: An Introduction* (Fifth Edition), 2007, by Roger Kirk, published by Wadsworth Publishing Company.
- *Statistical Reasoning in the Behavioral Sciences* (Fifth Edition), 2007, by Bruce King and Edward Minium, published by John Wiley & Sons.
- *Educational Assessment of Students* (Fifth Edition), 2006, by Anthony Nitko and Susan Brookhart, published by Prentice Hall.
- *Measurement and Assessment in Teaching* (Tenth Edition), 2008, by Robert Linn, Norman Gronlund, and David Miller, published by Prentice Hall.
- *Educational Measurement* (Fourth Edition), 2006, edited by Robert Brennan and sponsored by the American Council on Education and the National Council on Measurement in Education, published by Greenwood Publishing Company.

The following two educational measurement professional journals often contain articles and features that focus on many of the topics covered in this textbook. They tend to be at a fairly high technical level, but they do provide helpful perspectives on a number of important measurement topics.

- *Educational Measurement: Issues and Practice*, published on behalf of the National Council on Measurement in Education by Blackwell Publishing.
- *Applied Measurement in Education*, published by Routledge (Taylor and Francis, Inc.).

ADMISSION TESTING SELF-ASSESSMENT MODULE

This Self-Assessment Module is based primarily on the content of the Foundations of Admission Testing textbook for admission counseling professionals, supported by the publications listed in the Learning Assignments portion of each section. Questions are grouped under the titles of each corresponding section. This self-assessment is an “open-book” examination as those completing it may and should utilize the information available as it pertains to each question.

QUESTIONS

Section 1: Introduction

Question 1. The background for the work of the NACAC commission on the Use of Standardized Tests in Undergraduate Admission was based on the core observations of the National Research Council report, *Myths and Tradeoffs: The Role of Tests in Undergraduate Admissions*. Please list three of the clear benefits of tests that were listed in the *Myths and Tradeoffs* report:

1. _____
2. _____
3. _____

Question 2. The *Myths and Tradeoffs* report expressed concern that test scores were used in ways that were not in line with their design or stated purposes, were beyond their technical capacities or were detrimental to important widely shared goals of a fair, open and effective admission process. Given these tradeoffs, the National Research Council offered a series of admission test recommendations to postsecondary institutions. Please list three of these recommendations:

1. _____
2. _____
3. _____

Question 3. The commission emphasized that a one-size-fits-all approach for the use of standardized tests in undergraduate admission does not reflect the realities facing the nation’s many and varied colleges and universities. The commission focused on five critical issues (one described in each commission report chapter) with regard to testing that affect all institutions and admission professionals and listed them as recommendations (in both the Executive Summary and the Challenge to Stakeholders). Please list three of these critical issue recommendations:

1. _____
2. _____
3. _____