



Fairness on the SAT[®] Suite of Assessments

January 2021

About College Board

College Board is a mission-driven not-for-profit organization that connects students to college success and opportunity. Founded in 1900, College Board was created to expand access to higher education. Today, the membership association is made up of over 6,000 of the world's leading educational institutions and is dedicated to promoting excellence and equity in education. Each year, College Board helps more than seven million students prepare for a successful transition to college through programs and services in college readiness and college success—including the SAT® and the Advanced Placement® Program. The organization also serves the education community through research and advocacy on behalf of students, educators and schools.

For further information, visit collegeboard.org.

© 2021 College Board. College Board and the acorn logo are registered trademarks of College Board. PSAT/NMSQT is a registered trademark of the College Board and National Merit Scholarship Corporation. SAT®, PSAT™ 10, PSAT™ 8/9, AP®, AP Potential™, and Skills Insight™ are registered trademarks of College Board. Khan Academy® is a registered trademark of Khan Academy.

Suggested citation: College Board. 2021. *Fairness on the SAT Suite of Assessments*. New York: College Board.

Contents

Introduction	1
SAT Suite Constructs, Purposes, Uses, and Test-Taking Populations	3
Fairness of the SAT Suite	6
Test Construction	6
Test Design.....	6
Test Development.....	9
Operational Administration	12
Differential Validity and Prediction Analyses for the SAT with First-Year GPA (FYGPA).....	12
Score Reporting and Interpretation	13
Score Reporting	13
Skills Insight	15
Item Challenge Process.....	16
Practice	16
SAT Suite Question Bank	17
Accessibility	18
Universal Design.....	18
Accommodations and Supports	18
Conclusion	21
References	22

Introduction

College Board believes in providing all test takers with a fair opportunity to demonstrate their achievement on the assessments of the SAT® Suite: the SAT, PSAT/NMSQT®, PSAT™ 10, and PSAT™ 8/9 (College Board 2017a, 2017b). Conceptually, *fairness* can be defined in terms of both equitable treatment of all test takers during a test administration and equal measurement quality across subgroups and populations. Best practices as well as standards 3.1–3.5 of the *Standards for Educational and Psychological Testing*, jointly produced by the American Educational Research Association (AERA), the American Psychological Association (APA), and the National Council on Measurement in Education (NCME), call for test publishers to “minimize barriers to valid score interpretations for the widest possible range of individuals and relevant subgroups” (2014, 63). An assessment should be built in such a way that the constructs being assessed are measured equally for all intended test takers and test-taking subgroups, and it should be administered in a manner that is fair and equitable for all test takers, regardless of gender, race/ethnicity, and other characteristics.

To accomplish these goals, four aspects of fairness, identified by the *Standards*, should be addressed when developing and administering an assessment.

1. *Fairness in treatment during the testing process.* Fairness in treatment involves “maximiz[ing], to the extent possible, the opportunity for test takers to demonstrate their standing on the construct(s) the test is intended to measure” (51). The *Standards* note that test makers have traditionally tried to meet this goal through standardization of the testing process—that is, by ensuring that all students are given the same instructions, testing time, and the like—but that test makers also increasingly recognize that “sometimes flexibility is needed to provide essentially equivalent opportunities for some test takers” (51) when accommodations and supports in testing do not compromise the construct being measured (e.g., reading achievement).
2. *Fairness as lack of measurement bias.* Per the *Standards*, bias in a measurement itself or in the predictions made from it may occur when “characteristics of the test itself that are not related to the construct being measured, or the manner in which the test is used,” lead to “different meanings for scores earned by members of different identifiable subgroups” (51). Bias in this sense can play out as differential performance on items and/or tests by identified subgroups equally matched on the characteristic of interest (e.g., mathematics achievement) and/or in differential predictions (inferences) about such subgroups. It is the responsibility of test makers to identify and root out such construct-irrelevant factors when these factors advantage or disadvantage defined subgroups of test takers.
3. *Fairness in access to the construct(s) being measured.* The *Standards* define accessible testing situations as those that “enable all test takers in the intended population, to the extent feasible, to show their status on the target construct(s) without being unduly advantaged or disadvantaged by individual characteristics (e.g., characteristics related to age, disability, race/ethnicity, gender, or language) that are irrelevant to the construct(s) the test is intended to measure” (52). Accommodations and supports may take such forms as providing students with visual impairments access to large-print versions of text (when visual acuity is not the construct being measured) and avoiding the use of regional expressions in test questions intended for a national or international audience.

4. *Fairness as validity of individual test score interpretations for the intended uses.* The *Standards* indicate that test makers and users should attend to differences among individuals when interpreting test data and should not generalize about individuals from the performance of subgroups to which they belong. Given those considerations, “adaptation to individual characteristics and recognition of the heterogeneity within subgroups may be important to the validity of individual interpretations of test results in situations where the intent is to understand and respond to individual performance,” but test makers also have to consider whether such adaptations may, for particular purposes, “be inappropriate because they change the construct being measured, compromise the comparability of scores or use of norms, and/or unfairly advantage some individuals” (53–54).

College Board embraces the fairness expectations articulated by the *Standards* and the overarching goal of ensuring the maximal inclusiveness, representativeness, and accessibility of its SAT Suite test materials consistent with the constructs, purposes, and uses of the tests. Through its fairness-related documentation, processes, procedures, trainings, and other support materials, College Board strives to ensure that the tests of the SAT Suite

- are appropriate for and accessible to a national and international test-taking population of secondary students, and defined subgroups of that population, taking a medium- to high-stakes assessment of college and career readiness;
- neither advantage nor disadvantage individual test takers or defined population subgroups of test takers due to factors not related to the constructs (e.g., reading comprehension, mathematics achievement) being measured;
- are free of content or contexts likely to give offense, provoke a highly distracting emotional response, or otherwise inhibit test takers from demonstrating their best work on the tests;
- accurately and fairly portray the diverse peoples of the United States and the world and convey the widest possible range of ideas, perspectives, and experiences consistent with the tests’ designs;
- make test content as fully and as widely accessible to as many test takers as possible through design and development processes yielding materials consistent with the principles of universal design and through a range of accommodations and supports for test takers with particular needs, while, to the fullest extent possible, remaining faithful to the constructs being measured; and
- have clearly articulated purposes and uses for which they and their data should and should not be used and have clearly indicated populations for whom the tests are designed.

The following document provides an overview of several interrelated issues of fairness as they pertain to the SAT Suite assessments—specifically, how College Board ensures the fairness of its test content, makes that content as accessible as possible, and provides accommodations and supports for students needing them. Preceding this discussion is a concise description of the SAT Suite intended to provide context for evaluating College Board’s test fairness practices. Additional resources on these topics are identified throughout as well as at the end of this document.

SAT Suite Constructs, Purposes, Uses, and Test-Taking Populations

As the *Standards* make clear at numerous points, test fairness cannot be evaluated separately from an understanding of what a given test is purporting to measure (its *construct* or *constructs*), what purpose(s) and use(s) it is intended to serve, and who composes the test-taking population. Consideration of test construct is important because potential modifications to test content or delivery in the name of fairness should aim to eliminate or reduce artificial barriers to access while preserving, as much as possible, the essence of the knowledge, skills, and/or abilities being measured. In other words, fairness in testing is, to a large extent, about minimizing *construct-irrelevant* factors precluding test takers from demonstrating what they know and can do. Providing a student who is blind a test in braille is very likely to be a reasonable modification when the construct being measured is reading (because while most people read visually, the underlying construct is comprehension of textual information) but not when the construct being measured is visual acuity. An understanding of test purpose and appropriate use is important as well because a test designed for one use—say, placement into a program—may or may not be suitable for another use—such as diagnosing deficiencies in performance—and because purpose and use inform the validity of inferences that can be drawn about test takers from their performance on the test. It is also critical to understand the intended test-taking population, both in general and in terms of identified subgroups, so that test design and development can be guided to maximize accessibility for all test takers, test delivery can anticipate and accommodate special needs of individuals in the population without compromising the construct(s) being measured, and test materials can be evaluated in relation to their suitability for the population as a whole and its constituent subgroups.

With this in mind, we now turn to a brief discussion of the constructs, purposes, uses, and test-taking populations of the SAT Suite of Assessments. This overview offers context for the subsequent description of the process of ensuring fairness on the SAT Suite.

The SAT Suite of Assessments (“SAT Suite”) is College Board’s collective term for the SAT, PSAT/NMSQT, PSAT 10, and PSAT 8/9 testing programs.

- The SAT is typically administered to high school juniors and seniors. The test measures essential prerequisites for postsecondary readiness and success as determined through an extensive, ongoing research process.
- PSAT/NMSQT and PSAT 10 are typically administered to high school sophomores and juniors. PSAT/NMSQT is administered in the fall of each academic year, while PSAT 10 is administered in the spring. The PSAT/NMSQT and PSAT 10 tests are identical in format and content, but only PSAT/NMSQT serves as a qualifying test for the National Merit Scholarship Corporation’s annual scholarship competition. PSAT/NMSQT and PSAT 10 serve as opportunities to check in on students’ progress toward postsecondary readiness and to focus students’ preparation for post-high school study.

- PSAT 8/9 is typically administered to eighth and ninth graders and serves as a baseline for assessing students' readiness for college or workforce training.

In ways that make sense for different grade levels, the four tests measure the same knowledge and skills, making it easier for students, families, and educators to monitor student progress and address any areas of weakness.

Each test of the SAT Suite includes the following components:¹

- a Reading Test
- a Writing and Language Test
- a Math Test, including calculator and no-calculator portions

Each test is designed to collect evidence in support of a broad claim about student achievement.

- **Reading:** Students can demonstrate college and career readiness proficiency in reading and comprehending a broad range of high-quality, appropriately challenging literary and informational texts in the content areas of U.S. and world literature, history/social studies, and science.
- **Writing and Language:** Students can demonstrate college and career readiness proficiency in revising and editing a range of texts in a variety of content areas, both academic and career related, for expression of ideas and for conformity to the conventions of Standard English grammar, usage, and punctuation.
- **Math:** Students can demonstrate fluency with, understanding of, and the ability to apply the mathematical concepts, skills, and practices that are most strongly prerequisite and central to their ability to progress through a range of college courses, career trainings, and career opportunities.

The primary purpose of the SAT Suite assessments is to determine the degree to which students are ready to succeed in college and workforce training programs. All assessment content, which is developed in accordance with test designs grounded in the best available evidence about essential prerequisites for college and career readiness, aligns with this purpose. Because the SAT Suite assesses the content that research shows matters most for college and career readiness (e.g., College Board 2019), the resulting scores provide meaningful information about a student's likelihood of succeeding in postsecondary education. SAT Suite results should, however, not be used as the sole source of information for high-stakes decisions about students' academic achievement.

The SAT Suite assessments provide data that are used for many purposes by different users. The three key user groups are higher education officials, K–12 educators, and students and their families. In keeping with best practices and the requirements outlined in the *Standards* (AERA, APA, and NCME 2014), the SAT Suite's primary intended uses and the key interpretations for each group of primary users are discussed in the following paragraphs, with a rationale presented for each use.

Assessing students' college and career readiness (for use by K–12 educators, students, and students' families). The SAT College and Career Readiness Benchmarks ("SAT benchmarks") serve as challenging, meaningful, and actionable performance indicators

¹ The same fairness policies and processes also apply to the optional SAT Essay, which was available to all students through June 2021. Beginning with the 2021–22 academic year, the Essay will only be available where state contracts require it for federal/state accountability purposes.

of students' college and career readiness. States, districts, and schools use the SAT benchmarks to inform decisions about what proportion of their student body has a high likelihood of success in credit-bearing, college-entry coursework. Benchmark information is also provided to individual students. The SAT benchmarks are not intended for high-stakes decisions such as restricting student access to challenging coursework or discouraging aspirations of attaining higher education. Grade-level benchmarks are also provided through the PSAT-related assessments. These grade-level benchmarks indicate whether students are on track for college and career readiness and are based on expected student growth toward the SAT benchmarks at each grade.

Making college admission and course placement decisions (for use in higher education).

The SAT provides rich information on a student's level of preparedness for credit-bearing college-level work, helping admission professionals make more informed selection decisions. Colleges and universities are also able to use the detailed SAT score information, along with other data, to make more refined course placements for their students. This use is supported by predictive validity evidence examining the prediction of postsecondary outcomes as well as the accuracy of course placement decisions (Marini et al. 2019; Westrick et al. 2020).

Monitoring student progress through a vertically scaled suite of assessments (for use by K–12 educators, students, and students' families).

Every test in the SAT Suite is reported on a common vertical scale, with the SAT as the capstone measure. The SAT scales are established on a nationally representative college-bound population of juniors and seniors, while the scales for the PSAT/NMSQT and PSAT 10 tests and the PSAT 8/9 test are established on a nationally representative sample of students in grades 10 and 9, respectively. Establishing the scales in this manner allows for appropriate inferences of student growth and progress toward college and career readiness from year to year prior to a student taking the SAT. Statements about a student's level of preparedness for college and workforce training can then be made based on SAT performance. Students can track their own progress by using score information to identify knowledge and skill areas needing improvement and subsequently engage in practice opportunities that will help them become more prepared for postsecondary-level work.

Contributing to high school course placement decisions (for use by K–12 educators, students, and students' families). All assessments across the SAT Suite provide information about a student's readiness for particular AP[®] courses. AP Potential[™] results provide a strong indication of college readiness in a particular subject through actual student performance on the SAT, PSAT/NMSQT, PSAT 10, and PSAT 8/9 as well as information about the college-level classes for which they need to seek additional supports before enrollment.

Contributing to scholarship decisions (for use in higher education). Colleges and organizations that award scholarships, loans, and other types of financial aid to students may tie such assistance to students' academic qualifications, as reported by SAT Suite scores.

SAT Suite scores, however, should not be used as the single measure to rank or rate teachers, educational institutions, districts, or states, and users should exercise care when attempting to interpret test results for a purpose other than the intended purposes described here. College Board is not aware of any compelling validation evidence to support the use of any of the SAT Suite assessments, or other educational achievement measures, as the principal source of evidence for teacher or school leader evaluation. Assessment data, when properly used and subjected to several constraints, can, however, be used *in*

conjunction with other educational outcome measures to make inferences about school and educational quality, including teaching and learning.

The populations taking the SAT Suite assessments are inclusive of secondary-level students across grades 8–12. These students are both resident in the United States and, particularly for the SAT, around the world. In determining fairness policies and practices, College Board attends not only to these populations as wholes but also to identified subgroups within these populations. Consideration of such subgroups in test construction and administration is warranted by the AERA/APA/NCME *Standards* (2014) for a variety of reasons, including as part of efforts to eliminate measurement bias and to ensure equal access of all students and student subgroups to the test construct(s) being measured. In furtherance of these aims, College Board considers a number of population subgroups as part of qualitative and (when sample sizes permit) quantitative fairness analysis. These subgroups include (but are not necessarily limited to) male and female test takers and Black/African American, Asian American, Latino/a, Native American, multiracial, and White test takers.

Fairness of the SAT Suite

Test Construction

College Board has taken and continues to take numerous, exacting steps to establish the fairness of the SAT Suite. These efforts begin with the test design and its ongoing enactment through test development.

TEST DESIGN

Key Concepts

Fairness in testing begins with a thoughtfully crafted and sharply focused assessment design. Beginning in 2012, College Board undertook a thoroughgoing redesign of its flagship SAT and PSAT/NMSQT tests. The resultant SAT Suite—which also includes the PSAT 10 and PSAT 8/9—were purpose-built using the best available evidence in order to measure essential college and career readiness prerequisites. A central tenet of the design philosophy was that the tests would squarely address in greater-than-typical depth the relatively few topics that evidence shows matter most to readiness for and success in postsecondary undertakings. These topics include words and phrases as they are used in context; command of evidence; analytical writing in response to a source text; the math concepts, skills, and practices strongly associated with the requirements of a wide range of college majors, workforce training programs, and careers (“math that matters most”); problems grounded in real-world contexts; and analysis in science and in history/social studies. In addition, the design of assessment questions and problems was intended to more closely align than in the past with best practices in secondary teaching, thereby reducing the distance between assessment and instruction and making the test content more accessible and relevant to students.

Evidence Gathering and Consultation

In settling on these key elements and other aspects of the redesign of the SAT Suite assessments, College Board undertook an extended, intensive research and consultation

process. One aspect of this work, published in 2014 as part of the publicly available test specifications for the SAT, was a précis of research evidence in support of important aspects of the redesign (College Board 2014, section II). Another was a wide-ranging consultation process. Before the blueprints for the SAT Suite tests were finalized, College Board had solicited feedback from more than 250 enrollment leaders, representing a broad cross section of higher education institutions, through a multiyear advisory working group, one-on-one interviews, group meetings, and surveys. College Board also engaged with K–12 stakeholders, including a range of advisory committees, professional associations, chief state school officers, and state and district clients of the SAT and PSAT/NMSQT (College Board 2014, 14–16).

This prerelease process continues to be augmented by ongoing research as well as feedback from higher education and K–12. Among the most prominent elements are the following:

- *National curriculum surveys.* Approximately every three to four years, College Board undertakes and publishes the results from a survey of a nationally representative sample of middle school, high school, and postsecondary instructors. The primary purpose of the survey is to ascertain what knowledge and skills are deemed prerequisite for readiness to participate successfully in common first-year, credit-bearing postsecondary courses across a range of subjects. This survey serves as a check on whether the SAT Suite assessments measure (and only measure) what postsecondary faculty deem critical for incoming students to already know and be able to do. The survey also yields data on secondary-level instruction, which, as previously noted, the SAT Suite assessments are designed to align to as closely as possible. From these (and other) data, College Board makes periodic adjustments as warranted to the SAT Suite assessments to improve alignment to postsecondary faculty expectations and to better represent best practices in secondary teaching. The most recent report was released in 2019 (College Board 2019).
- *Evidence base enhancements.* In 2020 College Board published a substantially broadened and deepened evidence base in support of its SAT Suite English language arts/literacy assessments (College Board 2020). College Board is producing a condensed version of that work, which will include updated math-related evidence as well (College Board, forthcoming).
- *Cognitive lab evidence.* In 2020 College Board completed documenting the results of a cognitive lab study of select SAT items in its Evidence-Based Reading and Writing (ERW) and Math sections (College Board and HumRRO 2020). The main conclusion of the study was that the evidence collected from think-alouds from nearly one hundred SAT-age students validated College Board’s claim that the SAT elicits cognitively complex thinking from test takers.
- *Test review committees.* As described in detail below, College Board convenes groups of secondary and postsecondary educators from around the United States as independent experts to review its SAT Suite test materials for content soundness and fairness. Feedback from these groups is used to refine or remove problematic test material prior to use. In addition to providing actionable guidance about specific test materials, these committees offer College Board developers ongoing, vital connections to and information about teaching and learning as they are undertaken in classrooms throughout the United States—insight that shapes future test development at a macro level.

Content and Statistical Specifications

Content and statistical specifications operationalize the broad elements of a test’s design in actionable, repeatable, transparent detail. Content specifications describe such features as the subject matter and contexts to be included and the knowledge and skills to be measured. Statistical specifications define such parameters as how easy or difficult test questions and problems should be, what the distribution of item difficulty should be, and what constitutes adequate statistical discrimination between test takers of differing achievement levels. For the SAT Suite, content and statistical specifications were established after a careful research and consultation process and are periodically reconsidered and refined as part of the evidence-gathering process described previously.

Among their virtues, detailed specifications for test materials help further the goal of test fairness by ensuring that test content, regardless of when or by whom it was developed, meets the requirements of measuring the desired construct(s), aids in achieving the specified purpose(s) and use(s) of the test, is suitable for the identified testing population and its subgroups, and is highly consistent in substance. In combination with other steps, such as statistical equating (see below), the use of carefully articulated specifications helps ensure that test takers, regardless of the date on which they take the test or the particular set of test materials they receive, have highly comparable testing experiences and that their performance is not influenced by significant variability in test materials (e.g., one student receiving a series of easy Reading Test passages and another student receiving a series of hard passages). College Board has created and continues to maintain extensive documentation for use internally and by its partners in the development of stimulus materials (e.g., passages, informational graphics), questions, and problems and in the construction of test forms (full testing experiences). These development guides include discussion of general issues, such as the definition of the intended construct and identification of testing purposes, uses, and population, as well as detailed process guidelines and examples of effective implementation.

As feedback in various forms (e.g., student performance data, input from independent external reviewers, comments from higher education educators and officials, evidence base reviews) is received, refinements are occasionally made to content and statistical specifications and documentation for the purposes of improving the validity, reliability, transparency, and fairness of the assessments. When this happens, College Board communicates those changes to test takers and other stakeholders, including educators and state and district users of SAT Suite programs.

Test Review Guides

College Board maintains detailed content and fairness review guides for the independent experts it employs to evaluate its test materials. The content of these guides aligns closely with that of the internal stimulus material, question/problem, and form development guides that direct College Board staff in their work. The external-facing guides include information about the review process, expectations for reviewers, general content and fairness guidelines, and bulleted lists of specific considerations. These detailed guides promote fairness not only through explicit instructions about reviewing test content for fairness but also by virtue of standardizing and calibrating the review process, thereby ensuring that different reviewers or groups of reviewers approach the review task in similar ways. These guides are reviewed and updated as needed to reflect any refinements to the test design. Guide updates may also be made to reflect changes in the guidance itself in response to evolving circumstances and emerging understandings of fairness-related issues.

TEST DEVELOPMENT

Guided by detailed documentation and carefully defined processes undertaken by highly qualified subject matter and measurement experts, College Board’s test development process for the SAT Suite is designed to yield high-quality, valid, reliable, and fair assessments appropriate for the constructs, purposes, uses, and populations identified earlier. As part of the development process, College Board staff employ various means, both quantitative and qualitative, to ascertain and maintain the fairness of test materials.

External Fairness Review

All SAT Suite test materials are reviewed by external, independent reviewers who are asked to evaluate the materials for fairness. As a group, these reviewers are typically active classroom teachers drawn from across the nation, teach at the secondary and postsecondary levels, and are deeply familiar with diversity and inclusion best practices, the student population of interest (and its subgroups), and the nature and purposes of the assessments. Reviewers are individuals from diverse backgrounds, live and work in different regions of the United States, and teach at different levels (secondary, postsecondary) and types (e.g., rural, suburban, urban) of institutions. Reviews cover broad-based issues of fairness as well as specific matters respecting

- race/ethnicity (Blacks/African Americans, Asian Americans, Latinos/as, and Native Americans);
- gender;
- English learners; and
- students with disabilities.

Fairness reviewers are charged with helping ensure that test stimuli, questions, and problems are broadly accessible to the wide-ranging student population that takes the SAT Suite; do not advantage or disadvantage individual test takers or identified subgroups of test takers on factors unrelated to the constructs being assessed; and address topics and texts that are appropriate for the audience of secondary students and the occasion of medium- to high-stakes testing. In addition to employing their own professional judgment and expertise, fairness reviewers are directed to apply criteria developed by College Board for the SAT Suite. These criteria include both general considerations and those specific to elements of individual tests (e.g., literature passages in Reading). This fairness framework addresses the primary focus for qualitative fairness reviews; the concept of fairness; the testing purposes and constructs; the audience for and occasion of testing; topic selection; individual and group portrayal; individual and group identification; ethnocentrism; and language use. These guidelines are sometimes expressly modified for particular kinds of content. For example, reviewers examining Reading Test passages are informed that, in a limited exception to the general criteria on language use, a small number of foreign words and phrases, slang terms, dialect, and/or idiomatic expressions may be acceptable in passages selected from works of U.S. and world literature, provided that sufficient context to enable understanding is available, because such elements are an authentic part of the real-world texts being sampled for the test.

While the criteria overviewed above are intended to be the primary basis for qualitative fairness evaluations of test materials, College Board encourages reviewers to draw on their professional judgment and expertise in order to apply the criteria flexibly and contextually. Moreover, reviewers are invited to raise issues that may not fall neatly into any of the above

categories as part of the effort to ensure that all potential fairness issues receive thoughtful consideration.

Depending on the stage in the test development process, fairness reviews may be conducted exclusively through individual written comments or also via meetings held remotely. In the former case, which applies to materials that have not yet been pretested with students, College Board staff have the latitude to make a range of revisions based on feedback (though less latitude to modify stimulus materials drawn from previously published sources) or may, if flaws are significant and/or pervasive, decide to halt further development. In the latter case, which applies to materials that have already been pretested and are nearing operational use (i.e., use in testing for which students receive a score), reviewers are asked to provide comments in advance of the meeting; these comments are read and considered by College Board staff, who prepare potential responses, such as replacements or, less commonly, edits, for discussion at the meeting. College Board staff raise issues that were identified by reviewers as high priority, that were commented on by multiple reviewers (regardless of whether these issues were rated highly in importance by reviewers themselves), or that were not identified as high priority but nonetheless are potentially serious. Guided by College Board staff, reviewers talk over these issues, evaluate College Board–proposed remedies when applicable, and potentially introduce new issues (or change their minds on issues they previously identified). Whether comments are made only in writing or also shared in a meeting, College Board staff carefully assess all feedback, make decisions informed by best practices and expert consensus, and produce records of how particular issues were resolved.

Pretesting

All SAT Suite questions and problems are pretested on a motivated sample of test takers that resembles the population of interest and is sufficient in size to allow College Board to evaluate the materials statistically in terms of difficulty, to discern whether the questions and problems can effectively differentiate between lower- and higher-achieving test takers, and to ensure that test takers from different racial/ethnic and gender groups did not perform differentially beyond established thresholds (see “DIF Analysis,” below). The data from typically one thousand to three thousand test takers responding to each question or problem are used to evaluate student performance on multiple-choice and (Math only) student-produced response items. Once questions and problems have been pretested and the statistics associated with them have been computed, the materials are again reviewed prior to operational use by measurement and content specialists (including active classroom teachers at both the secondary and postsecondary levels) for content accuracy, clarity, construct relevance, fairness, difficulty, statistical discrimination, and other aspects of soundness and quality.

DIF Analysis

Analyses of differential item functioning, or DIF, are conducted at the pretest and operational stages of test development to identify questions and problems that may function differently for members of different population subgroups. It is important to note at the outset that DIF analysis is not based on the past test performance of various population subgroups, nor is DIF intended to remove questions and problems from use that members of certain population subgroups do well on. Rather, DIF analysis serves to call attention to certain questions and problems on which samples of certain population groups of equivalent achievement performed markedly differently.

The underlying assumption in such analyses is that all test takers demonstrating the same level of achievement in the content area should have similar chances of answering each question correctly, regardless of subgroup membership. DIF occurs when samples from different subgroups matched on achievement (i.e., two samples with similar scores on the assessment) differ notably in their performance on a specific test question or problem. The presence of DIF provides a statistical indication that a question or problem may function differently for individuals belonging to one subgroup from the way it functions for those belonging to another subgroup who are at the same achievement level. Questions and problems exhibiting DIF are divided into those showing low, medium, and high levels of DIF, with such designations based on established statistical thresholds. Those test questions and problems exhibiting high levels of DIF have a greater-than-normal chance of measuring factors irrelevant to an assessment (such as those related to culture).

DIF analyses begin by examining any differences in performance on individual questions or problems relative to two groups of comparable achievement, referred to as the *reference group* and the *focal group*. Questions and problems identified after pretest as exhibiting DIF over an established threshold are excluded from use, as the consequences to test development are relatively minor at this stage and are easily outweighed by the benefits of removing potentially problematic material at a comparatively early point in the development process. Questions and problems exhibiting such levels of DIF at the operational stage but that did not exhibit comparable DIF at the pretest stage are generally retained for use assuming that their operational statistics are otherwise unproblematic, a practice supported by the fact that all such test materials have already undergone (and passed through) fairness review and have not previously exhibited high DIF. Questions and problems exhibiting high DIF at the operational stage and for which there are no corresponding pretest DIF data (because population subgroup sample sizes were too small to permit some or all DIF analyses) are reviewed by a specially convened group of internal and external experts, who are charged with either confirming the fairness of each question or problem by establishing that construct-irrelevant factors do not plausibly account for the observed performance difference or, alternately, determining that one or more of the questions or problems are unfair in some way with respect to one or more subgroups and should not be scored.

Qualitative review is a critical complement to DIF analysis. The presence of DIF signals the possibility that a question or problem may be biased, but the results of DIF analyses alone do not determine whether a question or problem is unfair. That judgment must be made by experts evaluating the question or problem, taking into consideration the purpose(s) of the assessment, the appropriateness of the question or problem given the purpose(s) of the assessment (i.e., whether the knowledge or skill being tested falls within the test domain), and whether any construct-irrelevant factors are present in the question or problem. Feedback from experts on questions and problems flagged for DIF can be used to inform decisions about whether a given question or problem should be counted toward student test scores and, more broadly, can help test developers identify or avoid introducing construct-irrelevant factors in future development.

The value of qualitative review in DIF analysis is attested to by the AERA/APA/NCME *Standards* (2014). Although the *Standards'* discussions of DIF indicate that DIF procedures are a common and expected part of test analyses, the *Standards* also point out that statistical evidence of DIF does not necessarily imply a flaw or weakness in a question or problem. In fact, the *Standards* suggest that when DIF occurs, test developers should try to "identify plausible explanations for the differences" (82) and then *may* choose to remove

the question(s) or problem(s). More emphatically, the *Standards* assert that “there needs to be a suitable, substantial explanation for the DIF to justify the conclusion that the item is biased” (51). This stance warrants College Board’s policy of pairing qualitative review with quantitative and retaining some items that quantitative analysis alone would remove.

Equating

It is common practice in large-scale assessment programs to develop multiple test forms in order to support numerous test administrations over several years. This collection of test forms is purposefully developed to be parallel by building the forms according to the same test specifications and, on the basis of pretest statistics, making them comparable in terms of difficulty. However, exactly equal form difficulty can rarely be achieved in practice. As a result, the difficulty level even of test forms built to be parallel will vary to some degree. Equating is a statistical procedure that accounts for such differences in difficulty and is used to make reported scale scores interchangeable across multiple test forms (AERA, APA, and NCME 2014). Equating thus contributes to test fairness by ensuring that test scores carry the same meaning regardless of on which day or with which form students tested.

Operational Administration

Fairness also involves equality during test administration across all groups of test takers. For instance, detailed procedures are specified by College Board to ensure that each SAT Suite assessment is administered uniformly across all testing sites in a fair and equitable manner.² Security measures are also in place to ensure that no test taker or group of test takers obtains access to information or opportunities that allow them to attain scores by fraudulent means and thereby jeopardize the validity and fairness of the results of the assessment.

Differential Validity and Prediction Analyses for the SAT with First-Year GPA (FYGPA)

Fairness extends beyond question/problem performance and test construction; it is also strongly tied to test validity. AERA/APA/NCME Standards 3.6–3.8 (2014) address the notion that fair assessments ensure validity of test score interpretations across all subgroups within the test-taking population. For instance, the *Standards* assert that it is the responsibility of the test developer to ensure that test scores maintain the same meaning across various subgroups. In this case, test scores should not provide different criterion predictions for different subgroups. This evidence is obtained by investigating predictive validity by subgroup, or differential validity and prediction.

The first operational differential validity and prediction study of the redesigned SAT (summarized in Marini et al. 2019) was based on the college performance of more than 223,000 students from 171 four-year institutions. Findings were analyzed by gender, race/ethnicity, best language, and highest parental education level. A consistent finding from the differential validity analyses was that the relationship between the SAT and first-year GPA (FYGPA) is positive for all subgroups studied, with correlations ranging from .44 to .57. These are considered “medium” to “large” correlations (Cohen 1988). Similar to previous research, the study found that the SAT and high school GPA (HSGPA) tended to have slightly stronger predictive relationships with FYGPA for female students, Asian and White students, students with higher parental education levels, and students whose best language is English Only.

² Such uniformity, however, does not exclude the provision of appropriate accommodations and supports for students with particular needs. See “Accessibility” and “Accommodations and Supports,” below.

Moreover, the multiple correlations for the SAT and HSGPA with FYGPA always exceeded the correlations between HSGPA alone and FYGPA. The increases in the correlations indicate that the SAT adds incremental validity beyond HSGPA for all student subgroups when predicting FYGPA. More information is better than less.

Results from the differential prediction analyses in Marini et al. (2019) were also consistent with previous research. In terms of gender, male students' FYGPA tended to be slightly overpredicted (i.e., they performed slightly lower than predicted) by both the SAT and HSGPA, and female students' FYGPA tended to be slightly underpredicted (i.e., they performed slightly higher than predicted) by both the SAT and HSGPA. The SAT and HSGPA both tended to overpredict the FYGPA of students who identified as American Indian or Alaska Native, Black or African American, Hispanic or Latino, Native Hawaiian or Other Pacific Islander, and Two or More Races. Though the results varied across subgroups, in most cases the amount of overprediction or underprediction was the smallest when the SAT and HSGPA were used together.

Though there was some differential prediction for all subgroups, be it overprediction or underprediction, these differences were minor. When the SAT and HSGPA were used to predict FYGPA, the absolute values of the effect sizes—Cohen's d —were less than 0.20 for all but one small subgroup. These findings confirm that institutions should use multiple measures to predict FYGPA and that the amount of overprediction and underprediction will be minor.

Taken together, these differential validity and prediction study findings show that, in general, the utility of the SAT and its added informational value above HSGPA to predict FYGPA holds across the student subgroups examined. Institutions can feel confident using SAT scores and HSGPA for admission, scholarship, and advising/retention decisions across various student and institutional subgroups. The small differences found in the prediction of FYGPA by subgroup echo what has been found in previous research and in most cases are not even categorized as small effect sizes, but they are still important to study, document, and further examine.

Score Reporting and Interpretation

SCORE REPORTING

In keeping with the AERA/APA/NCME *Standards* (2014), SAT Suite score reports have been developed at the student and institutional levels to provide their intended audiences with appropriate interpretations of the reports and guidelines outlining the appropriate use of test results.

Student Score Reports

Online and paper score reports are available to students taking the SAT, PSAT/NMSQT, PSAT 10, and PSAT 8/9. Student score reports display scores on the scales of the SAT- or PSAT-related assessments, as applicable, including total, section, test, cross-test, and subscore scales, providing details to students on how they performed in specific areas (displayed in online and paper score reports). Score ranges are displayed online with every score, indicating how students would perform if they were to take the same test repeatedly with no new learning. Percentile ranks allow students to compare their scores to those earned by other students. The SAT College and Career Readiness Benchmarks are displayed at the section level (Evidence-Based Reading and Writing, Math) and are based on actual student success in entry-level college courses. The SAT benchmark scores represent a 75 percent

likelihood of a student achieving at least a C in a first-semester, credit-bearing college course in a related subject. The grade-level benchmarks for the PSAT-related assessments are based on expected student growth toward the SAT benchmarks at each grade. Whereas SAT benchmarks indicate likelihood of success in college, grade-level benchmarks indicate whether a student is on track for college and career readiness for their grade. Red, yellow, and green score indicators provide insightful feedback for reported scores, indicating specific areas of strength and weakness. Skills Insight™ information (College Board 2017c; discussed in more detail below) is designed to help students acquire a better understanding of how scores relate to specific academic knowledge and skills. Skills Insight offers descriptions of typical performance at each score band represented (e.g., 25–29 on the Reading Test, which, for the SAT, has a scale from 10 to 40). It also provides actionable suggestions for building knowledge and skills that help students gain additional practice (displayed in online score reports only).

K–12 Institutional Score Reports

Scores for the SAT, PSAT/NMSQT, PSAT 10, and PSAT 8/9 are released through an integrated score reporting portal (<https://collegereadiness.collegeboard.org/educators/highered/reporting-portal-help>), which supports effective decision-making with standard reports, interactive data analysis tools, and secure file downloading.

A number of institutional report types are available for the SAT Suite.

- **Scores by Institution** provides aggregate performance data by institution as well as student-level performance data at each school for every score.
- **Scores by Demographics** provides aggregate performance data by demographic group for every score. It also displays aggregate performance data for the entire institution and allows for the reporting of one or two demographic groups at a time.
- **Benchmarks by Institution** measures aggregate as well as student-level performance against College Board benchmarks for the Evidence-Based Reading and Writing section and the Math section.
- **Benchmarks by Demographics** provides aggregate performance against College Board benchmarks for the Evidence-Based Reading and Writing section and the Math section by demographic group. It also displays aggregate performance data for the entire institution and allows for the reporting of one or two demographic groups at a time.
- **Instructional Planning** highlights instructional strengths and weaknesses and can be used to help focus improvement efforts. It measures aggregate performance data against College Board benchmarks for every score except the total score and cross-test scores.
- **Question Analysis** provides performance data for every multiple-choice and student-produced response question and problem on the assessment. The performance data available are based on the type of test form—disclosed or nondisclosed—that was used for the administration. When a test administration uses a disclosed form, the Question Analysis report provides educators with aggregate-level reporting access with the following data: each test question/problem and answer explanation; correct answer; percentage of students who got the question correct by group (school, district, state, and total group); student responses by answer option (key and distractors) by group; difficulty level; and related cross-test scores and subscores. If an educator has K–12 Detail-level access, the Question Analysis report also enables the user to see student-level performance for each question/problem. When a test administration uses a nondisclosed test form, the Question Analysis report provides educators with aggregate-level reporting

access with the following data: percent correct by group (school, state, and total group); difficulty level; and related cross-test scores and subscores. If an educator has K–12 Detail-level access, the Question Analysis report also enables the user to see student-level performance for each question/problem. Comparisons to district, state, and total group are based on the data reporting hierarchies maintained by College Board.

- **Roster Summary** provides aggregate and student-level test administration information, including registration, fee waiver usage, absenteeism, and scores for each student registering for and/or completing the test.
- **Electronic Score Reports (Data Files)** are available to educators in the online score reporting portal's Download Center and are available in .txt and .csv file formats for integration into existing K–12 reporting systems (e.g., Naviance).

Designated Institution Score Reports

Students can opt to send their SAT scores to colleges, scholarship programs, and other designated institutions via official score reports. Colleges and scholarship administrators use SAT scores for admission applications and other opportunities.

Colleges can choose to receive paper or electronic score reports (data files). Electronic score data files are available to higher education institutions in the Higher Education Reporting Portal (<https://collegereadiness.collegeboard.org/educators/higher-ed/reporting-portal-help>).

The Higher Education Reporting Portal provides higher education institutions access to additional reports and features:

- The **SAT Trend Dashboard**, which provides year-over-year SAT score send trends
- **Executive Summary reports**, viewable by Enrollment Planning Service (EPS) subscribers

SKILLS INSIGHT

Skills Insight (College Board 2017c) provides sets of data-driven statements intended to facilitate interpretation of performance on the Reading, Writing and Language, and Math Tests of the SAT, PSAT/NMSQT, PSAT 10, and PSAT 8/9. Skills Insight statements were derived from in-depth analysis by College Board subject matter and measurement experts of the aggregate performance of millions of students on thousands of SAT Suite test questions and problems. The statements are organized by test score band (6–14, 15–19, 20–24, 25–29, 30–34, and 35–40). For each score band, the Skills Insight statements describe what a student scoring within that band is likely to know and be able to do in relation to the academic knowledge and skills measured on the tests. The goals of Skills Insight are to help students, teachers, administrators, and others understand what a test score means with respect to student achievement and, for scores below the highest range, how performance may be improved.

To independently validate the Skills Insight statements, College Board conducted an external review in late 2016. College Board recruited subject matter experts knowledgeable in the instruction of reading, writing and language, and math at the middle school, high school, and postsecondary levels. The selection process yielded a diverse group of experts in terms of gender, race/ethnicity, and geographic region within the United States. The external reviewers often had expertise at more than one instructional level. The subject matter experts were asked to consider whether the Skills Insight statements were clear, relevant to the curriculum, and aligned in a continuum from low to high levels of achievement across low to high score bands. The reviewers also examined whether the Skills Insight statements

in a particular range were consistent with the knowledge and skills needed to answer the ten to twenty sample test questions/problems assigned to each score band. Another task for the subject matter experts was to review the Skills Insight Suggestions for Improvement for clarity, appropriateness, and usefulness (i.e., whether the suggestions, if followed, would be likely to help students improve their academic achievement). Finally, the reviewers were asked to provide revisions to any of the statements and Suggestions for Improvement that would enhance the use of Skills Insight as an interpretative tool. The external reviewers found the Skills Insight statements to be well aligned with expectations for academic knowledge and skills in the score bands, the sample questions and problems corresponding to the score bands to be consistent with the Skills Insight statements for those bands, and the Suggestions for Improvement to be generally accurate and helpful guidance to students. External reviewer comments and specific revisions were thoughtfully evaluated by College Board content area staff as they further refined the SAT Suite Skills Insight material.

ITEM CHALLENGE PROCESS

The SAT Suite’s item challenge process offers additional transparency and a check on the soundness and fairness of test materials. Test takers may alert proctors to potential issues with test materials (which are filed with College Board via irregularity reports) or may contact College Board directly through established email channels. Test takers’ queries are routed to senior test development staff, who review the materials in question and, if applicable, develop answer explanations. If the test form on which the material appears is being released, test takers are supplied with details of the review’s findings; if the form is not being released, test takers are advised that the material was reviewed and what the outcome of the process was. In the rare circumstance in which a review would identify a substantive problem with the materials, College Board would undertake appropriate remediating action, up to and including removing a problematic question or problem from scoring.

Practice

A critical aspect of test fairness rooted in but extending beyond design, development, and administration is test practice. *Practice* includes the vital area of test familiarization—that is, making students, families, and educators aware of and comfortable with test instructions, formats, delivery methods, scoring, and the like—but more broadly encompasses resources and activities promoting the development of the durable knowledge and skills central to the construct(s) being assessed. Practice in this latter, expanded sense may be thought of schematically as a series of concentric circles. The innermost circles represent resources and activities that closely mirror the test itself and whose main purpose is to prepare test takers for the material they will encounter on test day. As one moves out from the innermost circles, resources and activities mirror the actual test and its requirements less and less directly but still retain their focus on the underlying knowledge and skills fundamental to its construct(s). In other words, while the resources and activities represented by the inner circles look like traditional “test prep,” those represented by the larger outer circles resemble those associated with the broader enterprise of teaching and learning—in fact, eventually becoming indistinguishable from teaching and learning. Test familiarization retains an important place in this scheme—test makers have a clear professional and ethical obligation to acquaint test takers and other stakeholders with their testing instruments—but practice confined to “test prep” risks narrowing the curriculum and mistaking the proxy of the knowledge and skill domains of interest (i.e., the test) with the domains themselves.

Practice, as described above, relates to test fairness in two critical ways. First, ensuring that all test takers have access to accurate, thorough information about the test well in advance of test day helps foster the goal of equity by giving everyone an equal chance to learn what is expected of them on the assessment, to address knowledge and skill gaps, and to avoid wasting valuable test time on reading directions, figuring out what the questions and problems are asking, and trying to understand how their responses will be scored. Second, a sustained emphasis on practice as the development of durable, broadly applicable knowledge and skills rather than as mere “test prep” holds the promise of helping to close achievement gaps and maintaining the full richness of the curriculum.

College Board provides a wealth of practice-related resources, most of them free of charge, to students and other stakeholders. These include (as of this writing) eight full-length SAT Practice Tests (<https://collegereadiness.collegeboard.org/sat/practice/full-length-practice-tests>; also available in print) and four full-length PSAT/NMSQT / PSAT 10 Practice Tests (<https://collegereadiness.collegeboard.org/psat-nmsqt-psat-10/practice/full-length-practice-tests>), including answer explanations; the *Official SAT Study Guide* book (MSRP of approximately \$30); and the full test specifications for the SAT (College Board 2014).

In addition, in keeping with its aim of providing to all test takers the same access to information and opportunities, College Board partners with Khan Academy® to provide free, personalized study resources for the SAT Suite assessments (<https://www.khanacademy.org/sat>). These include thousands of practice questions that adhere to guidelines established by College Board, multiple practice tests developed by College Board, and personalized recommendations for practice to help students fill their knowledge and skill gaps.

Studies of student SAT Suite test-taking performance have documented the efficacy of time spent on Official SAT Practice through Khan Academy. In two large-scale observational studies, students participating in Official SAT Practice for six or more hours exhibited notable score gains relative to their peers who did not use Official SAT Practice. In the larger and more recent study, Weatherholtz et al. (2020) determined that students who practiced with Khan Academy for at least six hours scored an average of 21 points higher on their first SAT than those students who did not. Moreover, those students using Official SAT Practice who also enacted at least one of three identified “best practices”—leveling up skills, taking a full-length practice test, and/or following personalized skill recommendations—roughly doubled that gain, to 39 points. These findings held when gender, race/ethnicity, and parental education level were controlled for.

SAT Suite Question Bank

The value of test familiarization to fairness extends also to making educators aware of and comfortable with the content on the SAT Suite assessments and to helping them prepare their students more effectively for test day. In support of these aims, College Board has created a resource called the SAT Suite Question Bank (SSQB) (<https://collegereadiness.collegeboard.org/educators/k-12/sat-suite-question-bank>). This free resource allows K–12 teachers to create custom, targeted question sets with the goal of improving instruction. The bank lets educators access over 3,500 questions from the SAT, PSAT/NMSQT and PSAT 10, and PSAT 8/9 assessments. Teachers can find questions that align with skills taught in class and use them in multiple ways. Questions can be sorted by the subscores, cross-test scores, and content dimensions they contribute to. Questions may

also be used as formative assessments that are short, simple, and achievable. In addition to its value in instruction, this bank contributes to the overall goal of SAT Suite test fairness by helping both teachers and students become more familiar with test content and formats and to see the profound connections between classroom teaching and learning and the durable knowledge and skills assessed by the SAT Suite.

Accessibility

College Board is strongly committed to making its SAT Suite test materials accessible to all test takers. This commitment extends both to making the assessments of the SAT Suite maximally accessible for all students and to ensuring that students with special needs receive appropriate accommodations and supports for test taking.

UNIVERSAL DESIGN

College Board subscribes to the principles of universal design, which, as noted by the *Standards*, has as its goal “develop[ing] tests that are as usable as possible for all test takers in the intended test population, regardless of characteristics such as gender, age, language background, culture, socioeconomic status, or disability” (AERA, APA, and NCME 2014, 57). To this end, College Board has taken a number of steps, including using a highly legible print layout, undertaking carefully structured studies to ensure that the time allotted for each test is adequate for the majority of test takers (thereby improving the likelihood that the test measures the intended construct rather than speed of test taking), and establishing guidelines for accessibility in the development and presentation of the informational graphics found on the Reading, Writing and Language, and Math Tests. Accessibility remains an ongoing area of focus for College Board, and refinements to the print and digital presentation and delivery of test materials will continue to be made as research and practice in the field identify ways to make tests more accessible to all.

ACCOMMODATIONS AND SUPPORTS

While observing the principles and adopting the practices of universal design and of accessibility more generally are helpful in reducing the number and severity of construct-irrelevant barriers for all test takers, some test takers may still need accommodations and supports in order to complete the assessment and/or obtain valid and fair test scores. The following section addresses the accommodations and supports routinely offered by College Board for the SAT Suite assessments.

Accommodations for Students with Disabilities

Testing programs address concerns about test validity and fairness for students with disabilities by offering test takers accommodations for testing that make available alternate formats or modify the manner in which tests are administered. The former includes such formats as large-print text, audio and screen-reader formats, and braille, while the latter includes extra breaks, extended time, modified test settings, and accommodations for recording responses. AERA/APA/NCME Standards 3.9 through 3.14 discuss the responsibility of test developers to develop and provide test accommodations and offer guidance on the appropriate use of said accommodations (2014, 67–70). These accommodations must be documented and also must allow for testing to be conducted without changing the construct(s) being measured, such that scores maintain their meaning across all subgroups as well as for both accommodated and nonaccommodated students.

An important way that College Board seeks to provide a fair testing environment for all test takers is by allowing students with disabilities to take the tests in the SAT Suite with the accommodations they need. This practice ensures that, when appropriate and possible, construct-irrelevant barriers that can interfere with test takers accurately demonstrating their true standing on a construct are removed (AERA, APA, and NCME 2014). As previously discussed, a *construct-irrelevant barrier* is any factor unrelated to the concepts or characteristics the assessment is designed to measure that can lead to an unfair testing experience and distort test takers' scores, decreasing the validity of the scores for their intended uses.

The accommodations offered by College Board serve to remove unfair disadvantages for those students with disabilities who have been approved to use accommodations on College Board assessments. In keeping with the *Standards* and best practices, accommodations are intended to respond "to specific individual characteristics but [do] so in a way that does not change the construct the test is measuring or the meaning of scores" (AERA, APA, and NCME 2014, 67). To this end, all accommodated forms and testing conditions are designed to be comparable to their nonaccommodated counterparts in that even though test forms or conditions might be modified based on the needs of a particular test taker, the construct being tested and the meaning of the score remain unchanged.

College Board's Services for Students with Disabilities (SSD) office authorizes a broad range of test accommodations. These accommodations include, but are not limited to, braille tests, large-print tests, and extended testing time. Students who show that their disabilities affect their ability to participate in College Board assessments are eligible for accommodations. The online request processing system, SSD Online (<https://accommodations.collegeboard.org/ssd-online>), allows schools to request accommodations and to track the progress of the request. Parents and guardians may request accommodations without the participation of schools via a paper request form. Once approved for accommodations by College Board, students are permitted those accommodations on the SAT, PSAT/NMSQT, and PSAT 10 assessments as well as on AP Exams.

The following examples of accommodations available from College Board illustrate many of the ways the organization helps eligible students receive the assistance they need. Some accommodations are administered in the standard testing room, as part of the administration. In select instances, an SSD coordinator administers accommodated tests in a separate space. Schools do not need to request approval from College Board to administer accommodations on the PSAT 8/9 assessments at this time, but all of the accommodations listed in the following text are available for that assessment as well. Please note the accommodations listed below are only examples; the list is not exhaustive.

Presentation

- Large-print (14 pt., 20 pt.)
- Human reader (Note: Reader reads entire test.)
- Sign language interpreter for verbal instructions
- Magnification device
- Prerecorded audio (MP3 via streaming)
- Colored overlays
- Braille

- Raised line drawings
- Assistive technology compatible (screen reader–accessible) format

Responding

- Writer/scribe
- Computer word processor for essays, with spellcheck, grammar check, and cut-and-paste features deactivated (AP only)
- Record answers in test booklet
- Large print (large block) answer sheet
- Use of four-function calculator in Math Test—No Calculator
- Braille device for written responses

Timing and Scheduling

- Extra or extended breaks
- Extended time
- Limited test time (may or may not include extra time)
- Specified time of day

Setting

- Small-group setting
- One-to-one testing
- Alternative testing site (with proctor present)
- Preferential seating
- Wheelchair accessibility
- Permissible food/drink/medication
- Permission to check blood sugar

Although numerous accommodations are possible, students with disabilities do not qualify automatically for SAT, PSAT/NMSQT, or PSAT 10 testing accommodations but instead must submit a request for approval by College Board. The vast majority of students who are approved for and are using testing accommodations at their school through a current Individualized Education Program (IEP) or 504 Plan have those same accommodations approved for College Board assessments. Most private school students with a current, formal school-based plan that meets College Board criteria also have their current accommodations approved for College Board assessments.

In those instances where a student does not qualify for automatic approval through the school verification process, the request and documentation are reviewed by College Board’s SSD Program. In general, students approved by SSD to receive College Board testing accommodations meet the following criteria:

- **The student has a documented disability.** Examples of disabilities include, but are not limited to, visual and hearing impairments, learning disorders, and medical impairments. Students must have documentation of their disability, such as a current psychoeducational evaluation or a report from a doctor. The type of documentation needed depends on the student’s disability and the accommodations being requested.

- **Participation in a College Board assessment is impacted.** The disability must result in a relevant functional limitation that impacts the student’s ability to participate in College Board assessments. For example, students whose disabilities result in functional limitations in reading, writing, and sitting for extended periods may need accommodations on College Board assessments, given the components of many of the tests and the manner in which assessments are generally administered.
- **The requested accommodation is needed.** The student must demonstrate the need for the specific accommodation requested. For example, students requesting extended time should have documentation showing that they have difficulty performing timed tasks, such as testing under timed conditions.

Approved accommodations remain in effect until one year after high school graduation (with some limited exceptions) and can be used on the SAT, PSAT/NMSQT, PSAT 10, and AP Exams. Students **do not need** to request accommodations from College Board for subsequent assessments taken during this eligibility period. As noted above, approval from College Board is not required for the use of accommodations on the PSAT 8/9 assessment at this time, and the accommodations available for other assessments may also be used on the PSAT 8/9. More information about the availability of accommodations and the procedures for requesting them prior to testing can be found on College Board’s SSD website, <https://www.collegeboard.org/students-with-disabilities>.

Supports for English Learners

To make the SAT even more accessible to students, College Board has worked with educators and state partners to provide testing supports for English learners (ELs). EL students taking a PSAT 10 or an SAT during the school day have access to testing instructions in several native languages as well as to approved word-to-word bilingual glossaries. These EL students do not have to apply to use word-to-word bilingual glossaries or translated instructions. Schools will be able to administer them directly to their students as needed, and students will receive a college-reportable score.

In addition, EL students may request extended time (time-and-a-half) on PSAT 10 or an SAT administered during the school day. To obtain this support, schools must visit the online request processing system mentioned above and select the EL Supports link.

Conclusion

As the above discussion shows, test fairness is a multidimensional concept that has important implications for nearly every aspect of test design, development, administration, reporting, and practice. The concept of fairness includes but extends far beyond requirements to ensure that individual test stimuli, questions, and problems are fair for all examinees. It also obligates test makers to define what they are measuring, for whom, and for what uses; to undertake exacting test design and development processes that promote the validity, reliability, and fairness of all test materials; to identify and evaluate subgroup performance differences and establish that test scores have the same meaning for members of different subgroups; to ensure that all students, including those with special needs, have the fullest possible access to the assessments; and to help ensure that all students are as

prepared as possible for test day, know and can interpret their scores, and have a clear sense of how to improve their performance as well as how to develop the underlying knowledge and skills being measured.

The preceding sections attest to the fact that College Board has an abiding commitment to test fairness in all its many forms for the SAT Suite of Assessments. In accordance with the AERA/APA/NCME *Standards* (2014) and best assessment practices more generally, and with a keen sense of its profound professional and ethical obligation to the students in its care and to the institutions and individuals who rely on and trust its data, College Board has undertaken numerous steps in the design, development, and administration of the SAT Suite assessments to achieve the goal of test fairness. As new needs arise and as better approaches are discovered, the organization will change apace, ensuring that its approaches meet or exceed the highest standards in the field.

References

American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. 2014. *Standards for Educational and Psychological Testing*. Washington, DC: American Educational Research Association.

Cohen, Jacob. 1988. *Statistical Power Analysis for the Behavior Sciences*, 2nd ed. Hillsdale, NJ: Erlbaum.

College Board. 2014. *Test Specifications for the Redesigned SAT*. New York: College Board. <https://collegereadiness.collegeboard.org/pdf/test-specifications-redesigned-sat-1.pdf>.

College Board. 2017a. *SAT Suite of Assessments Technical Manual: Characteristics of the SAT*. New York: College Board. <https://collegereadiness.collegeboard.org/pdf/sat-suite-assessments-technical-manual.pdf>.

College Board. 2017b. *SAT Suite of Assessments Technical Manual Appendixes*. New York: College Board. <https://collegereadiness.collegeboard.org/pdf/sat-suite-assessments-technical-manual-appendix-pt-1.pdf>.

College Board. 2017c. *Skills Insight for the SAT Suite*. New York: College Board. <https://collegereadiness.collegeboard.org/pdf/skills-insight-sat-suite.pdf>.

College Board. 2019. *College Board National Curriculum Survey Report 2019*. New York: College Board. <https://collegereadiness.collegeboard.org/pdf/national-curriculum-survey-report.pdf>.

College Board. 2020. *The SAT Suite and Classroom Practice: English Language Arts/Literacy*. Edited by Jim Patterson. New York: College Board. <https://collegereadiness.collegeboard.org/pdf/sat-suite-classroom-practice-english-language-arts-literacy.pdf>.

College Board. Forthcoming. *Evidentiary Foundations of the SAT Suite*. New York: College Board.

College Board and HumRRO. 2020. *The Complex Thinking Required by Select SAT Items: Evidence from Student Cognitive Interviews*. New York: College Board. <https://collegereadiness.collegeboard.org/pdf/sat-cognitive-lab-report.pdf>.

Marini, Jessica P., Paul A. Westrick, Linda Young, Helen Ng, Doron Shmueli, and Emily J. Shaw. 2019. *Differential Validity and Prediction of the SAT: Examining First-Year Grades and Retention to the Second Year*. College Board Research Report. New York: College Board. <https://files.eric.ed.gov/fulltext/ED597325.pdf>.

Weatherholtz, Kodi, Phillip Grimaldi, Catherine Hicks, Kelli Millwood Hill, Cassie Freeman, Bercem Akbayin-Sahin, Crystal Coker, Jennifer Ma, and Lara Henneman. 2020. *Use of Khan Academy Official SAT Practice and SAT Achievement: An Observational Study*. Mountain View, CA: Khan Academy.

Westrick, Paul A., Jessica P. Marini, Doron Shmueli, Linda Young, Emily J. Shaw, and Helen Ng. 2020. *Validity of SAT for Predicting First-Semester, Domain-Specific Grades*. College Board Research Report. New York: College Board. <https://research.collegeboard.org/pdf/validity-sat-predicting-first-semester-domain-specific-grades.pdf>.