SAT® Suite

2024

DIGITAL SAT SUITE OF ASSESSMENTS

# Technical Manual

## CHARACTERISTICS OF THE DIGITAL SAT SUITE

# Contributors

The following individuals were involved in the creation of the *Digital SAT Suite of Assessments Technical Manual*. We thank them for sharing so generously of their time, energy, and expertise.

## Chapter Leads

Angela DelBrocco, Jim Patterson, Thomas Proctor, Emily Shaw, Mark Syp

## Writers and Reviewers

Jim Patterson, Jen Bedet, Clancy Clawson, Garrett Ziegler, Tom Proctor, Oliver Zhang, Deanna Morgan, Hui Deng, Weiwei Cui, Jonathan Beard, Emily Shaw, Jessica Marini, Helen Ng, Paul Westrick, Linda Young, Sherral Miller, Laurie Morre, Kinny Mu, Beth Oxler, Angela DelBrocco, Abby Jacobs, Vincent Sasso, Carly Bonar, Dona Carling, Thom Gleiber, Sonia Wilson, Jenn Smith, Richard Green, Carolina Herrera, John Fallon; Tricia Renner, Yoko Kono, Ashleigh Goldberg, Cassandra Larson, Kathleen Weiss, Kathleen; Jay Happel, Christine Loew, Jill Green, Kevin Fay, Rajesh Rachabattuni, Josh Van Epps, Greg Bales, Jim Lauckhardt, Michael Polizzi, David Meade

## Technical Manual Team

Mark Syp, Chief Editor
Oliver Zhang, Group Lead
Len Carmichael, Quality Control Support

## Leadership Reviewers

Priscilla Rodriguez, College Readiness Assessments
Ana Menezes, Learning and Assessment
Kinny Mu, Assessment Design and Development
Sherral Miller, Assessment Design and Development
Julie Miles, Psychometrics

# Contents


**Contributors** ii

**Foreword** ix

**Preface** x

Chapter 1 **Overview of the Digital SAT Suite** 1
1.0 Introduction 1
1.1 Description and History of the SAT Suite 1
1.1.1 College Board and the SAT Suite 1
1.1.2 History of the SAT and the SAT Suite 3
1.2 The Digital SAT Suite 5
1.2.1 Overview 5
1.2.2 Transition to the Digital SAT Suite 8
1.2.3 Continuity and Change 11
1.3 Key Design Elements and Benefits of Digitization and MST 15
1.3.1 Key Design Elements 15
1.3.2 Benefits of the Digital SAT Suite 22

Chapter 2 **Test Specifications** 25
2.0 Introduction 25
2.1 Key Concepts 25
2.1.1 Purposes and Intended Uses and Interpretations 26
2.1.2 Constructs 29
2.1.3 Claims 30
2.2 Test Section Overviews 30
2.2.1 The Reading and Writing Section 30
2.2.2 The Math Section 36
2.2.3 SAT Essay 42
2.3 Statistical Specifications 46

Chapter 3 **Test Development and Assembly** 49
3.0 Introduction 49
3.1 Test Content Development 49
3.1.1 The Reading and Writing and Math Sections 49
3.1.2 The Digital SAT Essay 54
3.2 Inventory Management 57
3.3 Test Assembly 57
3.3.1 Multistage Adaptive Versions 57
3.3.2 Linear Versions 58

Chapter 4 **Test Administration and Security** 59
4.0 Introduction 59
4.1 Test Requirements 59
4.1.1 Administration 60
4.1.2 Test Length 60
4.1.3 Time Per Module 60
4.1.4 Total Number of Items 60
4.1.5 Total Time Allotted 60
4.1.6 Average Time Per Item 61
4.1.7 Scores Reported 61
4.1.8 Item Format(s) Used 61
4.1.9 Stimulus Subject Areas 61
4.1.10 Informational Graphics 62

4.2 **Digital Testing Application** 62
4.2.1 Digital Testing 62
4.2.2 Bluebook 63
4.2.3 Accessibility 64
4.3 **Form Assignment and Administration** 65
4.3.1 Multistage Adaptive Testing 65
4.3.2 Embedded Pretesting 67
4.3.3 Discrete Items 67
4.4 **Security Protocols and Analyses** 69
4.4.1 Test Security Rationale 69
4.4.2 Security Features 69

Chapter 5 **Test Scoring and Reporting** 71
5.0 **Introduction** 71
5.1 **Digital SAT Score Procedures** 71
5.1.1 Scoring Process 71
5.1.2 Scoring of the Digital SAT Essay 73
5.2 **Digital SAT Score Reporting and Resources** 75
5.2.1 Information About the Official Score Report PDF 77
5.2.2 Information About Knowledge and Skills 78

Chapter 6 **Psychometrics** 81
6.0 **Introduction** 81
6.1 **Adaptive Testing** 81
6.2 **Scaling and Norming** 83
6.2.1 Item Response Theory 83
6.2.2 Scaling 85
6.2.3 Normative Information 88
6.3 **Reliability and Errors of Measurement** 90
6.3.1 Simulation Studies 90
6.3.2 Standard Errors of Measurement 99
6.3.3 Reliability 100
6.3.4 Analysis of Score Structure 102
6.4 **Item Analysis, Calibrations, and Pre-Equating** 103
6.5 **Panel Assembly and Ongoing Psychometric Quality Management** 106
6.5.1 Panel Assembly 106
6.5.2 Ongoing Psychometric Quality Management 107

Chapter 7 **Fairness** 109
7.0 **Introduction** 109
7.1 **The Concept of Fairness in Testing** 109
7.2 **Fairness Reviews of Test Materials** 111
7.2.1 The Reading and Writing and Math Sections 111
7.2.2 The Digital SAT Essay 113
7.3 **Accessibility** 114
7.3.1 Universal Design 114
7.3.2 Universal Tools 119
7.3.3 Accommodations and Supports 119
7.4 **Test Administration** 123
7.4.1 Bluebook 123
7.4.2 Test Day Toolkit 123
7.5 **Test Security** 123

7.6 Additional Fairness Mechanisms ..... 124
7.6.1 Student Postexperience Surveys and Focus Groups ..... 124
7.6.2 Test Materials Challenge Procedure ..... 124
Chapter 8 **Validity** ..... 125
8.0 Introduction ..... 125
8.1 Introduction to Validity as a Concept ..... 126
8.2 Content-Oriented Validity Evidence ..... 127
8.2.1 Curriculum Survey ..... 127
8.2.2 Alignment to State College and Career Readiness Standards ..... 132
8.2.3 Subject Area Validity Evidence ..... 133
8.3 Response Process–Oriented Validity Evidence ..... 156
8.3.1 Cognitive Labs ..... 156
8.3.2 Usability/Accessibility Research ..... 158
8.4 Relationship Between the SAT and Other Educational Measures ..... 160
8.4.1 Sample ..... 160
8.4.2 Measures ..... 161
8.4.3 Methodology ..... 163
8.4.4 Results ..... 163
8.4.5 Summary ..... 165
8.5 Relationship Between the SAT and College Outcomes ..... 166
8.5.1 Methodology ..... 166
8.5.2 Measures ..... 169
8.5.3 Analysis ..... 171
8.5.4 Results ..... 171
8.5.5 Summary ..... 186
8.6 Measuring and Monitoring College and Career Readiness with the Digital SAT Suite ..... 187
Chapter 9 **Conclusion** ..... 189
9.0 References ..... 190
**Appendix A** ..... 204
**Appendix B** ..... 226

# Tables


Table 1.1 Digital SAT Suite Practice Opportunities ..... 19
Table 1.2 Digital SAT Suite Test Wisdom Resources ..... 20
Table 1.3 Digital SAT Suite Skill- and Knowledge-Building Resources ..... 21
Table 2.1 Digital SAT Suite Reading and Writing Section: High-Level Specifications ..... 32
Table 2.2 Digital SAT Suite Reading and Writing Section: Item Sequence ..... 33
Table 2.3 Digital SAT Suite Math Section: High-Level Specifications ..... 38
Table 2.4 Digital SAT Essay: High-Level Specifications ..... 43
Table 2.5 Digital SAT Essay: Scoring Dimensions and Evaluation Criteria ..... 44
Table 6.1 Bias of Estimated Theta and Scale Scores ..... 94
Table 6.2 RMSE of Estimated Theta and Scale Scores ..... 95
Table 6.3 Routing Decision Accuracy ..... 97
Table 6.4 Routing Decision Consistency ..... 97
Table 6.5 CCR Benchmark Classification Accuracy: Reading and Writing ..... 98
Table 6.6 CCR Benchmark Classification Accuracy: Math ..... 98
Table 6.7 CCR Benchmark Classification Consistency: Reading and Writing ..... 98
Table 6.8 CCR Benchmark Classification Consistency: Math ..... 99
Table 7.1 Digital SAT Suite Implementation of UDA Principles ..... 115
Table 8.1 Key Postsecondary Curriculum Survey Findings: Reading and Writing, All Surveyed Faculty ..... 128
Table 8.2 Key Postsecondary Curriculum Survey Findings: Math, Postsecondary Math Faculty ..... 129
Table 8.3 Digital SAT Suite Reading and Writing Text Complexity Ranges by Testing Program ..... 134
Table 8.4 Student Demographics of the Convergent Validity Study Sample ..... 161
Table 8.5 Descriptive Statistics for Measures of Interest, Overall Convergent Validity Study Sample ..... 162
Table 8.6 Correlations and 95% Confidence Intervals Between Digital SAT and Paper and Pencil SAT Scores (n=6,373) ..... 163
Table 8.7 Digital SAT and Paper and Pencil SAT Correlations with HSGPA, PSAT/NMSQT Total Score, and Average AP Exam Score ..... 165
Table 8.8 Institutional Characteristics of the Pilot Predictive Validity Study Sample ..... 167
Table 8.9 Student Characteristics of the Pilot Predictive Validity Study Sample ..... 168
Table 8.10 Descriptive Statistics for the Pilot Predictive Validity Study ..... 170
Table 8.11 Corrected (Raw) Correlation Matrix of Digital SAT Scores and HSGPA in the Pilot Predictive Validity Study ..... 171
Table 8.12 Corrected (Raw) Correlation of Predictors with Overall First-Year GPA (k=11, n=1,889) ..... 172
Table 8.13 Corrected (Raw) Correlation of Predictors with First-Year Credits Earned (k=11, n=1,889) ..... 183
Table 8.14 Digital SAT Suite College and Career Readiness and Grade-Level Benchmark Scores ..... 188

Table A-1 Digital SAT Suite Reading and Writing Section: Skill/Knowledge Testing Points ..... 204
Table A-2 Digital SAT Suite Reading and Writing Section: Operational Question Distribution ..... 206
Table A-3 Digital SAT Suite Math Section: Skill/Knowledge Testing Points—Algebra Content Domain ..... 207
Table A-4 Digital SAT Suite Math Section: Skill/Knowledge Testing Points—Advanced Math Content Domain ..... 211
Table A-5 Digital SAT Suite Math Section: Skill/Knowledge Testing Points—Problem-Solving and Data Analysis Content Domain ..... 215
Table A-6 Digital SAT Suite Math Section: Skill/Knowledge Testing Points—Geometry and Trigonometry / Geometry Content Domain ..... 218
Table A-7 Digital SAT Math Section: Operational Question Distribution by Content Domain ..... 220
Table A-8 Digital PSAT/NMSQT and PSAT 10 Math Section: Operational Question Distribution by Content Domain ..... 222
Table A-9 Digital PSAT 8/9 Math Section: Operational Question Distribution by Content Domain ..... 223
Table A-10 Digital SAT Suite Math Section: Operational Question Distribution Summary ..... 225
Table B-1 Institutional Characteristics of the Study Population for Sample Recruitment ..... 226
Table B-2 Student Characteristics of the Current Study Sample, 2020 SAT Validity Study Sample, and the 2022 College-Bound Seniors Population ..... 227

# Figure


Figure 3.1 Digital SAT Suite AIG Development Paradigm ..... 51
Figure 4.1 Digital SAT Suite MST Model ..... 66
Figure 6.1 Example Reading and Writing MST Panels ..... 82
Figure 6.2 Reading and Writing Section Cumulative Theta Distributions by Grade Level ..... 87
Figure 6.3 Math Section Cumulative Theta Distributions by Grade Level ..... 87
Figure 6.4 Grade 10 Math Section Cumulative Distribution by Assessment Level ..... 88
Figure 6.5 Conditional Bias by Difficulty Route: Reading and Writing ..... 91
Figure 6.6 Conditional Bias by Difficulty Route: Math ..... 92
Figure 6.7 Conditional Scale Score Bias: Reading and Writing ..... 94
Figure 6.8 Conditional Scale Score Bias: Math ..... 94
Figure 6.9 Conditional RMSE: Reading and Writing ..... 95
Figure 6.10 Conditional RMSE: Math ..... 96
Figure 8.1 Digital SAT and Paper and Pencil SAT Correlations with HSGPA, PSAT/NMSQT Total Score, and Average AP Exam Score ..... 164
Figure 8.2 Mean First-Year GPA by Digital SAT Total Score Band ..... 173
Figure 8.3 Mean First-Year GPA by Paper and Pencil SAT and Digital SAT Total Score Band ..... 174
Figure 8.4 Mean First-Year GPA by Digital SAT Total Score Band: Subgroup Analyses ..... 175
Figure 8.5 Mean First-Year GPA by Digital SAT Total Score Band within HSGPA for “A” Students ..... 176
Figure 8.6 Probability of Earning a First-Year GPA of 3.00 or Higher, by SAT Total Score and HSGPA ..... 177
Figure 8.7 Mean First-Year Math GPA by Digital SAT Math Section Score Band ..... 178
Figure 8.8 Mean First-Year STEM GPA by Digital SAT Math Section Score Band ..... 178
Figure 8.9 Mean First-Year All-But-Math GPA by Digital SAT Reading and Writing Section Score Band ..... 179
Figure 8.10 Probability of Earning a Math FYGPA of 3.00 or Higher, by Digital SAT Math Section Score and HSGPA ..... 180
Figure 8.11 Probability of Earning a STEM FYGPA of 3.00 or Higher, by Digital SAT Math Section Score and HSGPA ..... 180
Figure 8.12 Probability of Earning an All-But-Math FYGPA of 3.00 or Higher, by Digital SAT Reading and Writing Section Score and HSGPA ..... 180
Figure 8.13 Mean First-Year GPA for STEM Majors, by Digital SAT Total Score Band Within HSGPA ..... 181
Figure 8.14 Probability of STEM Majors Earning a STEM FYGPA of 3.00 or Higher, by Digital SAT Total Score and HSGPA ..... 182
Figure 8.15 Mean First-Year Credits by Digital SAT Total Score Band ..... 184
Figure 8.16 Mean First-Year Credits Earned by Digital SAT Total Score Band: Subgroup Analyses ..... 185
Figure 8.17 Probability of Earning 30 or More Credits in the First Year of College, by Digital SAT Total Score and HSGPA ..... 186

# Foreword

For over a century, the SAT® has provided students the opportunity to show colleges what they have learned in the core areas of reading, writing, and math. During this time, the test has undergone updates, revisions, and reworkings to better meet the evolving needs of students from all backgrounds and the always-changing landscape of K–12 and postsecondary education. As recently as 2015, the test was completely redesigned with an eye toward better measuring the knowledge and skills that students are learning in high school and that matter most for college and career readiness. It was then relaunched together with—for the first time—free, world-class personalized practice for all students.

Now the test has evolved once more, as the SAT Suite has embraced a fully digital testing experience. College Board's goal is to provide a less stressful, more accessible experience for all students and educators while retaining the value, rigor, and predictive power of the paper-based SAT Suite. To achieve this goal, we assembled a team of experts in math and literacy content development, assessment design, psychometrics, technology, and product development and rigorously tested and piloted these new digital assessments with students and educators around the world. In close consultation with College Board's K–12 and higher education members as well as experts in the field, this team has produced an all-digital suite of assessments that is easier to take, easier to give, more secure, and more relevant.

While the adaptive design of the digital test allows for a shorter, more secure, and more flexible assessment than paper and pencil, it provides the same proven assessment of student knowledge, with the content domains, constructs, and score scales remaining consistent. It is perhaps most important to recognize that the test remains technically sound, effectively serving its stated purposes and providing well-documented evidence of its psychometric properties.

In keeping with the best practices of the digital SAT Suite of Assessments launch, this manual documents the rationale, content, processes, and outcomes of the new digital SAT Suite, all presented with an emphasis on their relevance and the benefits they will bring to students. This manual also fulfills the crucial task of representing a baseline of evidence supporting the test development and psychometric quality of the SAT Suite.

Much like the development of the test itself, the creation of this manual involved the contributions of many College Board team members, all of whom were committed to presenting this documentation in an easy-to-read format, with clear, concise evidence supporting the stated uses of the assessments. It is our hope that this manual serves as an important resource to all who use and interpret the digital SAT Suite. Given the evolving and iterative nature of testing, the manual will continue to be supplemented as the SAT Suite is administered to more and more students nationally and internationally, and as we continue to develop and refine our assessments and how they can be most effectively used to promote student readiness based on the results of our ongoing research.

Priscilla Rodriguez

Senior Vice President, College Readiness Assessments

College Board

# Preface

## Purpose of This Manual

The purpose of the *Digital SAT Suite of Assessments Technical Manual* is to provide higher-education professionals, K–12 educators, students and their families, state education leaders, and others who use or who are interested in using the digital SAT Suite of Assessments with detailed, up-to-date information about the technical qualities of the suite and its assessments. This manual describes the purposes and intended uses of the assessments and the rationale and principles undergirding them. It also addresses the content of the tests; the procedures and processes undertaken in the creation, administration, and scoring of the tests; how test scores should be interpreted and used; how College Board ensures the accuracy of test scores; and the evidence College Board has collected that establishes the validity, reliability, and fairness of interpretations made on the basis of their scores.

College Board believes it is essential to provide documentation of this nature, in keeping with the organization's commitment to transparency and the need to adhere to industry best practices and professional standards, such as those outlined in the *Standards for Educational and Psychological Testing* by the American Educational Research Association, American Psychological Association, and National Council on Measurement in Education (AERA et al., 2014). As gathering and maintaining validity evidence supporting the SAT Suite is an ongoing, iterative process, this manual will be updated as additional information becomes available.

## Manual Contents

For ease of reading and understanding, the *Digital SAT Suite of Assessments Technical Manual* is structured in a manner broadly matching that of the lifecycle of one of the tests in the suite. The manual offers insights into the digital SAT Suite from its conception and design, to the development of its tests, to the tests' administration and scoring, and finally to the valid interpretations of those scores for intended uses.

As its name implies, Chapter 1, Overview of the Digital SAT Suite, provides an orientation to the digital SAT Suite of Assessments, including a general description of the suite and its history, information about what changed and what stayed the same in the transition from paper and pencil testing to digital, and a discussion of the key design features of the tests and the benefits offered by the digital testing model developed for them.

**Chapter 2, Test Specifications,** provides an overview of the content and statistical specifications of the digital SAT Suite. It includes information on the intended purposes and uses of the digital SAT Suite tests, definitions of the concepts of *constructs* and *claims* as they apply to assessments, overviews of the content of each test section, and information about the key measurement properties of the tests in the suite.

**Chapter 3, Test Development and Assembly,** details the processes and procedures used in the creation and delivery of the tests of the digital SAT Suite. It provides an overview of the production of test *items* (questions, problems, and tasks) and the rigorous approach College Board uses to develop and vet them, discusses how the created items are stored and maintained in College Board's item bank system, and describes how items are assembled into multistage adaptive and linear tests.

**Chapter 4, Test Administration and Security,** documents how College Board administers the digital SAT Suite in a manner that ensures that all test scores are valid for their intended purposes and uses, all test takers have a fair testing experience, and no secure test information is disclosed or shared. It begins by describing the parameters and procedures used to standardize the test administration process. It then discusses Bluebook™, the test delivery application used for the digital SAT Suite tests, and the benefits afforded to students by Bluebook, including its basis in universal design principles, its provision of universal tools to all students, and its support for a broad array of accommodations and supports for students who require them for fair access to the tests and their content. It continues with a closer examination of the components of the digital suite's adaptive testing model, and concludes with a discussion of the critical issue of test security and the measures College Board takes to ensure that the tests of the digital SAT Suite are secure.

**Chapter 5, Test Scoring and Reporting,** covers the scoring procedures employed for the digital SAT Suite tests and the means by which those scores are reported, including the various resources available to students and educators as ways to better understand, interpret, and make productive use of those scores.

**Chapter 6, Psychometrics,** discusses crucial aspects of measurement science as applied to the digital SAT Suite tests. Topics covered include a technical overview of multistage adaptive testing as used for the digital SAT Suite; how College Board established the reported scale and the process used to maintain reported normative information; the procedures College Board employs to ensure that the digital adaptive tests produce reliable test scores and consistently route test takers to an optimal test experience; the methods College Board uses to analyze pretest items prior to their operational use and the pre-equating procedures used to ensure comparable scores across operational tests; the procedures College Board follows to monitor the reported score scale and ensure that item parameters used to produce scores remain stable and accurate; and an overview of the test security analytics conducted by College Board as part of the broader effort to establish and maintain validity.

The manual then looks at the concepts of fairness and validity. **Chapter 7, Fairness,** discusses College Board's commitment to and the processes in place to ensure the fairness of all facets of the digital SAT Suite. It includes an overview of the concept of fairness in testing (particularly those relevant to the digital SAT Suite) and gives considerable attention to how this multifaceted concept informs and is implemented in the digital SAT Suite tests. It also details the rigorous internal and external review processes employed by College Board to ensure that the tests are fair to all students. It then turns to the critical issue of test accessibility and how maximal accessibility for all test takers is attained through the application of universal design principles, the provision of universal tools, and the availability of accommodations and supports. It next discusses how the concept of fairness is applied to test administration as well as the advances that the transition to digital adaptive testing in Bluebook as the primary delivery mode has made in this area. It concludes with an analysis of how security in the creation, handling, and delivery of test materials contributes to both the reality and perception of the fairness of the digital SAT Suite tests.

Finally **Chapter 8, Validity,** offers a wealth of evidence from a wide range of academic and empirically derived sources establishing the validity of the digital SAT Suite tests and the ways in which the assessments' scores can be used to evaluate college and career readiness and success—the overall goal of the suite.

College Board

May 2024

Chapter 1

# Overview of the Digital SAT Suite

## 1.0 Introduction

This chapter provides a broad overview of the digital SAT Suite of Assessments. Section 1.1, Description and History of the Digital Suite, provides a general description of the SAT Suite and its history as a way to contextualize the significance of College Board's move in 2023 and 2024 from paper and pencil to digital adaptive testing as the suite's primary delivery mode. Section 1.2, The Digital SAT Suite, provides an overview of the digital suite as well as information about what changed and what stayed the same in the transition from paper and pencil to digital adaptive testing. Section 1.3, Key Design Elements and Benefits of Digitization and MST, lays out the key design features of the digital SAT Suite tests as well as the benefits offered by the suite's use of multistage adaptive testing (MST).

## 1.1 Description and History of the SAT Suite

### 1.1.1 College Board and the SAT Suite

College Board is a mission-driven not-for-profit organization that connects students to college and career 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 six thousand of the world's leading educational institutions and is dedicated to promoting excellence and equity in education. College Board reaches more than 7 million students a year, helping students navigate the path from high school to college and career through its programs—including the SAT, the Advanced Placement® (AP®) Program, and BigFuture®. The organization also serves the education community through research and advocacy on behalf of students, educators, and schools.

The *Digital SAT Suite of Assessments* is College Board's collective term for its flagship suite of digital college and career readiness testing programs and services: the SAT, PSAT/NMSQT® and PSAT™ 10, and PSAT™ 8/9. The digital suite represents an evolution of the SAT Suite that debuted in the 2015–2016 academic year. While continuing to measure the skills and knowledge assessed by the paper-based SAT Suite it replaced, the digital suite is responsive to the changing educational landscape as well as the emerging needs of students and their families, teachers, state and district users, higher education officials, and policymakers. Over the several years that the SAT Suite was available in its redesigned paper-based form, College Board listened closely to feedback and input from a wide range of stakeholders, carefully assessed the needs of the suite's users, and evaluated how best to respond. The result is the digital SAT Suite.

Like the previous paper-based SAT Suite, all assessments of the digital SAT Suite measure English language arts/literacy and math skills and knowledge. To this end, each digital test contains a Reading and Writing section and a Math section. The digital SAT (only) may also be accompanied by the digital SAT Essay, a direct-writing assessment administered as part of select U.S. school day administrations at the request of particular state education departments.

The digital SAT, College Board's primary college and career readiness assessment, is a key component of the digital SAT Suite, with the PSAT/NMSQT, PSAT 10, and PSAT 8/9 tests serving as grade-appropriate assessment options for middle school/junior high school and high school students.

For nearly a century, the SAT has been used successfully worldwide, in combination with factors such as high school grade point average (HSGPA), to assess student preparedness for and to predict student success in postsecondary education. In the graduating class of 2023, more than 1.9 million test takers took the SAT (College Board, 2023a, SAT Participation and Performance table), the results of which were used by thousands of high school educators and postsecondary admission officers around the world. Beyond admissions and college readiness, the SAT has several other uses and interpretations, including monitoring student growth, contributing to course placement decisions, connecting students to career and other opportunities, helping historically underrepresented students be seen by colleges and universities during the recruitment process, and connecting students to scholarships.

The digital SAT Suite is administered via Bluebook, College Board's digital testing application, which administers the tests in an intuitive, accessible manner. The digital SAT was first administered at international test centers in spring 2023. Starting in fall 2023, all PSAT-related testing, both domestic and international, moved to digital. In spring 2024, all students took the SAT digitally. Except for those students who require a paper-based test form (e.g., as an accommodation), the shift to digital testing is now complete.[1]

The digital suite continues and expands on the paper-based suite's core commitments to access and opportunity for all students. These commitments include:

- Offering valid, reliable, fair, and objective assessments of students' academic achievement
- Providing actionable information to students and educators about evidence-based ways to build on academic strengths and to address skill and knowledge shortcomings relevant to college and career readiness
- Connecting students to opportunities they have earned through their hard work in school, such as admission to postsecondary institutions well suited to their achievement and interests as well as scholarships and recognitions
- Helping state users meet federal accountability requirements through industry-leading assessments, services, and documentation
- Helping higher education institutions to find and enroll prospective students and then to support those students so that they can be successful on their campuses.

1 Beginning in 2018, College Board made linear (fixed-form, nonadaptive) digital versions of several of the paper-based SAT Suite tests available to state and district users who wanted to administer the exams via computer. Those versions of the SAT Suite assessments were digitized versions of the paper-based tests, with small modifications to improve the user experience, and were retired alongside the paper-based suite. This document's references to the "paper and pencil SAT Suite" or "paper-based SAT Suite" include these linear digital versions as well.

## 1.1.2 History of the SAT and the SAT Suite

In keeping with standards and best practices in large-scale standardized assessment (AERA et al., 2014), the SAT has been reconfigured several times over its nearly one-hundred-year history, with each iteration addressing evolving circumstances and making the test sounder, fairer, and more useful to students and their families, teachers, counselors, college admission staff, and other stakeholders.

In 1926, 8,040 young men took what was then called the Scholastic Aptitude Test at its first administration. This initial version of the SAT bears little resemblance in design and structure to the current test of the same name. The former was explicitly grounded in a paradigm of aptitude testing, which seeks to identify (in this case) academic potential rather than measure the acquisition of specific skills and knowledge obtained through coursework, with the intent of helping colleges find students who had the capacity (aptitude) for postsecondary education irrespective of whether they had received traditional college-preparatory instruction. It contained nine subtests, seven with verbal content and two with mathematical content. Beginning in 1930, the SAT was split into two sections, one portion designed to measure verbal aptitude and the other to measure mathematical aptitude.

In 1959, College Board created the PSAT (then called the Preliminary Scholastic Aptitude Test) to provide students with an inexpensive way to prepare for the SAT. In the years since, the PSAT has grown into its own student-focused assessment program, serving as an opportunity to check in on student progress toward college and career readiness. In 1971, the PSAT became the Preliminary Scholastic Aptitude Test/National Merit Scholarship Qualifying Test (PSAT/NMSQT). In addition to its traditional role, the PSAT/NMSQT served (and continues to serve) as a qualifying test for the National Merit® Scholarship Corporation's annual scholarship program.

In 2005, the "new" SAT debuted. This substantially revised test had three sections (critical reading, writing, math) instead of the traditional two, was about three-and-a-half hours in length, and was scored on a 600–2400 scale; replaced analogy items with short passage-based reading comprehension items; added a required direct-writing assessment alongside multiple-choice revision and editing items; and began testing some elements of Algebra II instruction along with skills and knowledge typically developed in the first two years of high school instruction.

In the fall of 2008, College Board field-tested the ReadiStep assessment, which was designed to measure the skills and knowledge necessary for eighth and ninth graders to be considered on track for college and career readiness. This assessment became the starting point of College Board's College and Career Readiness Pathway, which also included the PSAT/NMSQT and the SAT. ReadiStep's content was closely aligned with that of both the PSAT/NMSQT and the SAT. The PSAT 8/9, which replaced ReadiStep in the College Board Readiness and Success System, is administered in the eighth and/or ninth grades. For students, the PSAT 8/9 is the earliest opportunity they have to engage with the SAT Suite of Assessments and serves to establish a baseline for assessing their college and career readiness.

The redesigned SAT Suite was launched in the 2015–2016 academic year. Drawing on emerging trends as well as input from many sources over the intervening years since the debut of the "new" SAT in 2005, the redesigned suite introduced eight important changes to testing in the suite:

- The elimination of the testing of obscure vocabulary (“SAT words”) in favor of assessing students’ acquisition and use of high-utility academic words and phrases in context
- The inclusion of items requiring students to demonstrate command of evidence, both textual and quantitative
- The introduction of an optional source-based direct-writing assessment, which replaced the required essay of the “new” SAT and whose use, alongside other changes, meant that the SAT would return to its traditional 400–1600 scale
- An emphasis on the math that matters most for high school and postsecondary success
- The inclusion of literacy and math items grounded in real-world contexts, including careers
- Sustained attention to the application of analytical skills in history/social studies and science contexts
- Direct attention to U.S. founding documents and texts in the Great Global Conversation
- A move to rights-only scoring, which eliminated the guessing penalty imposed in previous iterations of the test

Structurally, the redesigned SAT Suite tests consisted of two required sections—(1) Evidence-Based Reading and Writing and (2) Math—and an optional Essay at the SAT level (only). The Evidence-Based Reading and Writing section consisted of two tests—(1) Reading and (2) Writing and Language—while the Math section consisted of the Math Test. The optional digital SAT Essay, initially offered to all students and later only as part of select U.S. school day administrations, yielded three dimension scores (Reading, Analysis, and Writing), which were not combined with each other or with any other SAT scores. The optional Essay replaced the generic writing fluency task of the “new” SAT with a source-based rhetorical analysis task, and students were given double the time (50 minutes instead of 25) offered for the “new” SAT Essay to respond to the task.

In the spring of 2016, College Board launched the PSAT 10 test as part of the redesigned suite. The PSAT 10 is the same test as the PSAT/NMSQT but is delivered in the spring rather than the fall of a given school year and does not serve as a qualifying test for the National Merit Scholarship Corporation’s annual scholarship program. The test’s purpose is to provide schools, districts, and states with a suite-based option for testing students’ college and career readiness progress in the spring of their sophomore year.

In the 2010–2011 academic year, College Board began offering the SAT during the school day to eligible U.S. district and state partners. The SAT School Day program enables participating states and districts to create an opportunity for all their juniors or seniors to take the SAT in their home schools during normal instructional time. SAT School Day provides encouragement for all students to pursue a college education and offers improved access and convenience to meet college admission testing requirements. School day administrations are also offered for the PSAT/NMSQT in addition to fall Saturday testing. The PSAT 8/9 and PSAT 10 are available only as school-based administrations.

The digital SAT Suite is the most recent iteration of the evolving set of assessment programs and services. In the redesign of the paper-based SAT Suite implemented during the 2015–2016 academic year, College Board carefully examined what the best available evidence indicated were the essential prerequisites in reading, writing, and math for readiness for and success in college and careers. This evidence, along with extensive feedback from colleagues in K–12 and higher education, was critical to shaping the design of the digital suite of assessments delivered today.

## 1.2 The Digital SAT Suite

### 1.2.1 Overview

The digital SAT Suite is a series of testing programs and related services designed to measure students' attainment of what the best available evidence has identified as essential college and career readiness outcomes in English language arts/literacy and math.

This section describes the four digital SAT Suite testing programs, their purposes, and the uses and interpretations intended for them and their data, with the goal of informing readers about the place of the digital SAT Suite and its assessments in the broader educational landscape.

#### Testing Programs

Like its paper and pencil predecessor, the digital SAT Suite consists of four testing programs, each with its own purpose and target population:

- 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 and is used in college admissions around the world.
- **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 program. PSAT/NMSQT and PSAT 10 serve as opportunities to check in on students' progress toward postsecondary readiness, focus students' preparation for post–high school study, and connect students to scholarship opportunities and College Board's National Recognition Program.
- **PSAT 8/9** is typically administered to eighth and ninth graders and serves as a baseline for assessing students' readiness for college and career.

The four tests measure the same broad knowledge domains and skills, with slight modifications reflecting differences in the age and attainment of students across the secondary grades, allowing students, families, and educators to monitor student progress and address any areas in need of improvement.

## Purposes and Intended Uses and Interpretations

The primary purpose of the digital SAT Suite is to determine the degree to which students are prepared to succeed both in college and careers. All assessment content, which has been developed based on high-quality research identifying the knowledge and skills most essential to college and career readiness and success, aligns with this core purpose (for details see Chapter 2, Test Specifications, and Chapter 3, Test Development and Assembly). Each test within the digital SAT Suite is designed to collect evidence from student performance in support of a set of broad claims about what students know and can do, and each claim is aligned to the primary purpose of assessing college and career readiness. The resulting scores provide meaningful information about a student's likelihood of succeeding in college and workforce training—information that, used in conjunction with other data (such as high school grades) and in the context of where a student lives and learns, can contribute to decisions about higher education admission and placement.

Although the core purpose of the digital SAT Suite is college and career readiness assessment, the suite's data are employed for many purposes by a range of users, notably higher education officials, K–12 educators, and students. In keeping with best practices and professional standards (AERA et al., 2014), the digital SAT Suite's intended uses and interpretations are discussed in greater detail, along with a rationale for each use, in Chapter 2, Test Specifications.

**Evaluating and monitoring students' college and career readiness** (for use by K–12 educators and students). The digital SAT's empirically derived College and Career Readiness Benchmarks ("SAT benchmarks") serve as challenging, meaningful, and actionable indicators of students' college and career readiness. Grade-level benchmarks established for the digital PSAT-related assessments 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. Additionally, the PSAT 8/9 and SAT are used to satisfy federal accountability requirements in 8th and 11th grades in several states.

**Monitoring student progress through a vertically scaled suite of assessments** (for use by K–12 educators and students). Every test in the digital SAT Suite is reported on the same vertical scale, with the digital SAT as the capstone measure. Having a single vertical scale allows for appropriate inferences regarding a student's academic growth and their progress toward college and career readiness from year to year prior to them taking the digital SAT. One is then able to make statements about a student's level of preparedness for college and careers based on digital SAT performance.

Contributing to high school course placement decisions (for use by K–12 educators and students). All assessments across the SAT Suite provide information about a student's readiness for particular Advanced Placement (AP) courses. AP Potential™ results provide a more challenging indication of college readiness in a particular subject through actual student performance on the SAT, PSAT/NMSQT, PSAT 10, and PSAT 8/9, and the AP Exams. These results can provide students with information about what college-level classes they are ready for in high school and courses for which they need to seek additional supports before enrolling.

**Connecting students to career possibilities** (for use by students). Discovering career options is a driving force as students make decisions about their future. Every career requires a set of skills, the attainment of which can be measured. College Board has worked with experts in occupations and labor market data to map the reading, writing, and math skills and knowledge measured on the SAT and the PSAT-related assessments to the literacy and numeracy requirements of a thousand different careers. To help all students consider the full range of vocational options open to them, digital SAT Suite score reports include the Career Insights Snapshot, which lists careers in a student's state that are connected to the student's assessment performance. Each listed career has a bright outlook, pays a living wage in the state, and requires some form and level of postsecondary education. These careers are presented as examples and are neither formal recommendations nor the only career options that students should consider.

**Connecting students to postsecondary educational opportunities** (for use by students). Connections™ is free and exclusively for students who take the PSAT/NMSQT, PSAT 10, or SAT on a school day. Connections lets students hear from nonprofit accredited colleges and universities (domestic and international), nonprofit scholarship providers, and government agencies administering educational programs. It delivers relevant messages via the BigFuture School app and by mail from organizations interested in students based on information that students, schools, districts, or states provide as a part of in-school assessments; score ranges on SAT, PSAT/NMSQT, PSAT 10, and AP Exams; as well as student interests and preferences shared in BigFuture School. Districts and schools may opt not to provide access to Connections. With Connections, no personal information is shared with institutions unless a student chooses to connect and share directly with them. This gives students and their families more control over when, or whether, they raise their hands to be seen. The goal is to create more opportunities for students as they consider their options after high school. With input from students, families, and education professionals, BigFuture School and Connections will expand and improve to help every student chart their path.

**Helping underrepresented students be seen by colleges** (for use by higher education). The College Board National Recognition Program awards academic honors to high-performing underrepresented students. The five national recognition programs include National First-Generation Recognition Program, the National African American Recognition Program, the National Hispanic Recognition Program, the National Indigenous Recognition Program, and the National Rural and Small Town Recognition Program. Students who take eligible administrations of the PSAT/NMSQT, PSAT 10, or AP Exams and meet the score requirements are considered for the awards, which are a tangible way to help students be seen by colleges and support colleges' recruitment strategies.

**Making college admission, advising, and college course placement decisions** (for use by higher education). The digital SAT provides information on a student's level of preparedness for college-level work, which helps admission professionals make more informed enrollment decisions. Once students are admitted, digital SAT results offer useful and valid ways to support students on campus, including informing course placement and academic major decisions and helping identify students in need of academic support.

**Contributing to scholarships and other awards** (for use by higher education and nonprofit organizations). Scores from the digital SAT Suite are often used to inform the decisions that colleges and nonprofit programs make in relation to academic awards, scholarships, and other forms of aid.

### Limits on Uses and Interpretations

The digital SAT Suite is intended to open doors for students and to help them gain access to opportunities that they have earned through their hard work. It is therefore inappropriate to use digital SAT Suite scores as a veto on students' educational or vocational aspirations. When interpreted properly, data from tests such as those of the digital SAT Suite can make valuable contributions to helping students meet their academic and career goals, but test scores should never be the sole basis for highly consequential decisions about students' futures. Digital SAT Suite scores, therefore, should be considered alongside other factors, including high school grades and where students live and learn, when evaluating students' achievement or potential.

Digital SAT Suite scores should also not be used as the single measure to rank or rate teachers, educational institutions, districts, or states. Users should exercise care when attempting to interpret test results for a purpose other than the intended purposes described above. College Board is not aware of any compelling validation evidence to support the use of any of the digital SAT Suite assessments, or other educational achievement measures, as the principal source of evidence for teacher or school leader evaluation. Assessment data, when 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.

For further examples of uses of College Board test scores that should be avoided, see *Guidelines on the Uses of College Board Test Scores and Related Data* (College Board, 2018).

## 1.2.2 Transition to the Digital SAT Suite

Among the many challenges that emerged or became more prominent in the educational landscape in the years since the redesign of the SAT Suite in the 2015–2016 academic year, four interrelated ones had a particularly important role in motivating College Board to design and implement the digital SAT Suite:

- Widespread, persistent concerns about the amount of time U.S. students spend taking tests
- Continued and growing threats to test security
- Ongoing concerns about the value and affordability of higher education
- The continued lack of college and career readiness attainment by a large proportion of students, especially those from historically underserved populations

The following subsections discuss each of these challenges in turn and indicate how the digital SAT Suite has been responsive to those challenges.

## Time Spent Testing

Since at least the No Child Left Behind Act of 2001 was signed into law, critics of standardized testing as well as many families, educators, and policymakers have raised concerns about the extent to which U.S. students are tested as part of K–12 education. Polling has suggested that the public's doubts about the value of standardized testing in schools have grown over time, and the necessary relaxation of federal testing requirements under the successor Every Student Succeeds Act during 2019–2020 and, to a lesser extent, the 2020–2021 pandemic years further contributed to those doubts. (For a brief recent overview, see Bruno & Goldhaber, 2021.)

The digital SAT Suite has responded to these concerns in two main ways. First, the digital-suite tests, like their paper and pencil predecessors, are useful and meaningful to students. The digital SAT Suite tests offer students the opportunity to evaluate their attainment of or progress toward achieving college and career readiness, and, as discussed in the subsection on higher education affordability and value below, they open doors and connect students to opportunities that they have earned through their hard work in school. Second, the digital SAT Suite exams are substantially shorter than their paper-based predecessors—about two hours in length rather than nearly three—and the move to digital testing has afforded schools and students more flexibility in when the tests are given.

## Test Security

Although the SAT Suite tests can open doors for students and connect them to opportunities they might otherwise miss, it can only do so if the tests themselves are secure and the results are accurate reflections of students' own efforts. Test security challenges, which are infrequent but highly consequential, threaten the integrity of the tests and the confidence that test takers and data users have in them. Over the long term, these threats, if unmet, erode trust in the tests. In the nearer term, they risk curtailing students' access to testing, as they can have a potential impact on scoring, up to and including rare situations in which scores or whole administrations are canceled due to security compromises.

A key motivation behind College Board's introduction of the digital SAT Suite was to meet these security challenges head-on and to do so in a way that expanded, rather than restricted, access to the tests. The digital-suite tests have reduced test security risks in a number of important ways, notably by eliminating the need to physically deliver, handle, store, distribute, collect, and reship paper test materials around the world, and by ensuring that each student who takes one of the digital tests is administered a highly comparable but unique version of the test.

## Higher Education Affordability and Value

In a further development in no small way attributable to the pandemic but indicative of wider concerns about the cost and value of higher education, enrollment in postsecondary education has been on the decline in recent years and has experienced only modest recent growth (National Student Clearinghouse Research Center, 2022, 2023, 2024). Although College Board (2023b) has observed a decadelong decline in inflation-adjusted tuition and fees at public two- and four-year institutions (alongside an increase at private nonprofit four-year colleges), higher education affordability remains a major concern for current and prospective students (Klebs et al., 2021) and for adults generally (Fishman et al., 2021). Indeed, respondents in both pre-pandemic surveys (e.g., Gallup & Lumina Foundation, 2015; Kaplan, 2015) and more recent ones (e.g., Citizens Financial Group, 2021; Fishman et al., 2021; Snyder, 2022) have cited college affordability as a major concern and stressor.

An obvious contributor to concerns about college affordability and value is worry about student debt. Statistics compiled by Hanson (2024) indicate that the average federal student loan debt balance is $37,088 and may, in fact, be closer to $39,981 once private loan debt is incorporated. As with many things in education, this burden is not borne equally by all. Hanson (2023a, 2023b) notes, among many other sobering findings, that 66% of all student loan debt belongs to women; that Black/African American holders of bachelor's degrees owe an average of $52,000 in student loan debt; and that Black/African American and Asian student borrowers have the highest monthly loan payments.

Negative perceptions of higher education affordability and value, and the troubling realities behind them, are important because they threaten the historic upward trend of college enrollment, which, in turn, is significant because, even with affordability being a concern, higher education retains tremendous value for both individuals and society. Reporting the results of College Board–led research, Ma and Pender (2023) reached four main conclusions about the benefits of higher levels of education:

- Individuals with higher levels of education earn more, pay more taxes, and are more likely than others to be employed.
- Median earnings increase with level of education, but there is considerable variation in earnings at each level of educational attainment.
- College education reduces the chance that adults will rely on public assistance.
- Adults with higher levels of education are more active citizens than others and are more involved in their children's activities. Having a college degree is associated with a healthier lifestyle, potentially reducing health care costs (pp. 4–5).

The digital SAT Suite helps promote these benefits by connecting students more easily and effectively than ever before to the opportunities they have earned.

- Students who take the PSAT/NMSQT, PSAT 10, and SAT as part of the school day can access BigFuture School, a free mobile app that gives them the power to plan for the future. In the app, students may receive customized career information and guidance about planning and paying for college and can opt in to Connections (where available), College Board's privacy-forward way for students to hear from nonprofit colleges, educational organizations, and scholarships that might be a good match for them without sharing any personal information.
- Students who take the PSAT/NMSQT test in the fall of their junior year can qualify for hundreds of millions of dollars in scholarships from the National Merit Scholarship Program and other partner organizations.
- Any student can access College Board's BigFuture (**bigfuture.org**), a free online guide to help all students take the right first step after high school. Regardless of a student's intended path after high school, the resources and tools on BigFuture help students explore careers, plan for college, and pay for college.
- Students who take a weekend administration can opt into College Board's Student Search Service™ (**studentsearch.collegeboard.org**), which helps colleges and scholarship programs reach out to students who may be a good fit based on their recruitment and selection criteria. All students have the option to join Student Search Service through their personal College Board account.

## College and Career Readiness Gaps

College and career readiness for all students by no later than the end of high school remains an essential but elusive goal, particularly for members of historically underserved population groups. The achievement picture has remained frustratingly steady, unacceptably low, and reflective of differential impact across the student population as well as societal and educational inequities. Among high school graduates in the class of 2023 who took the SAT, only 40% were considered college and career ready by meeting both of the empirically established College and Career Readiness Benchmarks (a reading and writing section score of 480 and a math section score of 530); 35% of these same graduates met neither benchmark (College Board, 2023a, SAT Participation and Performance: Total and Race/Ethnicity tables). Sadly, this reflects a broader trend, as the proportions of students in the class of 2023 who met one or more college and career readiness benchmark scores on the ACT® assessment were also low: 51% in English (down from 59% for students in the class of 2019), 30% in math (down from 39%), 40% in reading (down from 45%), and 31% in science (down from 36%), with only 21% meeting all four benchmarks (down from 26%) (ACT, 2023, p. 8, Table 1.1).

What is more, although 74% of Asian American students, 51% of White students, and 51% of students identifying as two or more races met both SAT benchmarks, only 24% of Hispanic/Latino students, 21% of Native Hawaiian/Other Pacific Islander students, 17% of American Indian/Alaska Native students, and 17% of Black/African American students did the same; among these latter groups, between 50% and 60% of students met neither SAT benchmark (College Board, 2023a, SAT Participation and Performance: Total and Race/Ethnicity tables).

Although college and career readiness testing cannot by itself eliminate these inequities, it plays a critical role in calling attention to and measuring progress toward closing these gaps. Given how significant, persistent, and consequential these gaps are, it would be unwise to turn away from the instruments that inform about them. This is not to say, however, that those instruments cannot be improved. Better tests—ones that are easier to take, easier to give, more secure, and more relevant to all students—can improve the test-taking experience; yield valid, reliable, actionable data; and clear pathways to opportunities. These features of better tests are hallmarks of the digital SAT Suite and are discussed in detail in the next section.

## 1.2.3 Continuity and Change

The digital SAT Suite represents both continuity and change with respect to the SAT Suite first administered in the 2015–2016 academic year. In essence, the digital SAT Suite is a refined evolution of the paper-based SAT Suite. At the domain level, the digital-suite assessments address content highly comparable to that found in the paper and pencil tests and retain strong alignment to essential college and career readiness prerequisites and, consequently, to state college and career readiness standards.

Change between the suites is primarily reflected in the move to digital and adaptive test delivery, substantially reduced test length, and modifications in test item format, particularly evident in the assessment of reading and writing skills and knowledge. The result is a set of assessments preserving the strong foundations of the paper-based suite while introducing innovations in flexibility, efficiency, focus, relevance, and security that make the digital-suite tests responsive to the educational moment and the needs of users.

This section begins with a discussion of the elements that have carried over to the digital suite from the paper and pencil suite and then continues with an overview of the changes introduced into the SAT Suite by the shift to digital testing.

## What Has Stayed the Same

The digital SAT Suite retains and builds on many of the key emphases of the paper-based SAT Suite. The digital SAT Suite continues to:

- Measure the skills and knowledge that students are learning in school and that matter most for college and career readiness
- Be scored on the same scales as the paper and pencil tests they have replaced, meaning that, for example, the SAT continues to be scored on the familiar 400–1600 scale
- Allow students and educators to track growth via an integrated, vertically scaled suite of assessments from grade 8 through high school and a series of empirically derived benchmark scores aligned with college and career readiness requirements
- Be administered in schools and test centers with a proctor
- Support students' readiness for test day and their development of relevant knowledge and skills through free, world-class practice resources, including Official Digital SAT Prep on Khan Academy®
- Connect students to scholarships
- Recognize the strong academic performance of underrepresented students through the College Board National Recognition Program
- Support all students who need accommodations and/or supports to access the tests and their content

The selected-response sections of all digital SAT Suite tests—(1) Reading and Writing and (2) Math—as well as the digital SAT Essay (offered only as part of select U.S. school day administrations) also demonstrate strong continuity with their paper and pencil predecessors. Key elements carried over from the paper-based suite include:

- The use of reading/writing passages across a range of academic disciplines and text complexities
- Required demonstrations of command of evidence, both textual and quantitative
- An emphasis on high-utility academic (tier two) words and phrases in context
- A focus on the revision and editing of writing to improve the effectiveness of expression, achieve specified rhetorical goals, and demonstrate command of core conventions of Standard English sentence structure, usage, and punctuation
- Continued stress on the math that matters most for college and career readiness and success
- Math problems in context as well as without context
- The use of both multiple-choice and student-produced response item formats in the Math section

- The continued availability of the SAT Essay's direct assessment of reading, analysis, and writing skills and knowledge as an optional part of select U.S. school day administrations, now in a digital format

Both because of this strong similarity in the content being measured and the fact that the two suites are grounded in the best available evidence about critical prerequisites for college and career readiness, the digital SAT Suite, like the paper and pencil suite, is strongly aligned to both postsecondary entry requirements and to state academic standards. The digital assessments also retain the key psychometric properties of the paper-based exams that users have come to expect and rely on from College Board.

Rigor is one of those properties. Although College Board has taken pains to make the experience of taking the digital SAT Suite tests easier than taking their paper and pencil predecessors, these efforts should *not* be confused with making the tests themselves easier. The tests continue to measure students' mastery of the knowledge and skills required to be ready for college and workforce training. As these requirements are challenging to attain, so must be the tests that assess their attainment.

The digital-suite tests, in other words, maintain the same level of challenge that the SAT Suite assessments have long been known for. Efforts to ensure comparable levels of rigor have taken a number of forms. Among the most notable are the following:

- College Board has aligned the digital-suite tests, like their paper-based predecessors, with the best available evidence about essential college and career readiness prerequisites.
- College Board continues to work closely with a range of independent experts, including subject matter experts at the secondary and postsecondary levels, to ensure that the tests and their items are sufficiently challenging to assess the knowledge and higher-order skills students need to be ready for college and careers.
- College Board employs robust content development and psychometric processes to verify that digital-suite test items are comparable in difficulty to those used on the paper and pencil versions of the tests.

In sum, while the digital SAT Suite assessments greatly simplify the test-taking process and give students better opportunities to show what they know and can do, the standards to which students are being held have not changed.

## What Has Changed

While preserving the best of the paper and pencil SAT Suite assessments, College Board seized the opportunity the transition to digital testing offered to reconsider and refine what was tested and how, all in the service of better meeting the needs of students and their families, educators, policymakers, and other stakeholders. These changes are reflected both at the suite level and in the test sections that compose the suite's assessments.

At the suite level:

- The digital SAT Suite assessments are substantially shorter than their paper and pencil predecessors—about two hours instead of three (exclusive of the optional digital SAT Essay).

- Test takers have more time, on average, to answer each Reading and Writing and Math question, meaning that, more so than ever before, the digital SAT Suite exams are measures of students' skills and knowledge, not test-taking speed.
- Students and educators receive scores faster than was possible with the paper and pencil SAT Suite.
- In addition to the many ways that the paper-based SAT Suite connected students to opportunities they had earned through their hard work, digital SAT Suite score reports include the Career Insights Snapshot, a list of growing careers in the student's state that connect to their scores. Career Insights Snapshot helps students consider career options and the postsecondary pathways needed to reach their goals. These careers are presented as examples and are neither formal recommendations nor the only career options that students should consider.
- The tests are more secure. Instead of large groups of students taking the same paper and pencil test form at the same time, each student taking one of the digital SAT Suite assessments is administered a highly comparable but unique version of the test. (How this is achieved is discussed more fully in Chapter 3, Section 3.3, Test Assembly.
- As a result of the increase in test security, states, schools, and districts have much more flexibility in terms of when they give the digital SAT Suite tests, including wider testing windows for the PSAT/NMSQT and SAT School Day.

At the test section level:

**Reading and Writing**

- The digital-suite assessments have a single Reading and Writing section instead of separate Reading and Writing and Language tests. This shift serves to make English language arts/literacy assessment on the digital SAT Suite tests more efficient while also acknowledging the reciprocal, mutually reinforcing nature of reading and writing skills and knowledge.
- The Reading and Writing section's passages are significantly shorter and more numerous, giving students more, and more varied, opportunities to demonstrate what they know and can do and to encounter information, ideas, and perspectives they find interesting and relevant. At the same time, these shorter passages maintain the level of rigor of longer reading passages with respect to text complexity and grounding in academic disciplines.
- A single (discrete) question is associated with each passage (or passage pair) instead of having several questions associated with a small number of longer passages, as was the case in the paper and pencil SAT Suite tests. For information on how the switch to discrete items benefits both students and the quality of the assessments, see Chapter 4, Test Administration and Security.

**Math**

- Calculators are allowed throughout the Math section. A single Math section has replaced the separately timed no-calculator and calculator-allowed portions of the paper and pencil SAT Suite Math tests. This change allows the Math section to more accurately reflect how the tool of the calculator is used in schools and in the real world. It also eases test administration by eliminating separately timed test portions with different rules. Students may continue to use their own approved calculator on test day or take advantage of the Desmos® Graphing Calculator, which is built directly into Bluebook.

- The average length, in words, of in-context items ("word problems") has been reduced. In-context items still serve a valuable role in the Math section, as they assess whether students can apply their math skills and knowledge to both academic and real-world situations. However, College Board listened to feedback that longer contexts posed barriers that could inhibit some students, often but not only English learners, from demonstrating their core math achievement.

# 1.3 Key Design Elements and Benefits of Digitization and MST

## 1.3.1 Key Design Elements

A number of key design elements characterize the tests of the digital SAT Suite. These include:

- Digital testing as the primary test delivery method (with paper-based and other accommodations and supports for students who require them)
- Bluebook, College Board's digital testing application built to administer the digital SAT Suite in an intuitive and accessible manner
- Multistage adaptive testing, which permits shorter tests that nonetheless yield scores as precise and reliable as those from the paper-based SAT Suite tests
- Embedded pretesting, which ensures that College Board can securely obtain high-quality item performance statistics and maintain the digital SAT Suite indefinitely while limiting the burden on students of answering pretest (nonoperational) items on which they are not scored
- The use of discrete items to assess skills and knowledge in English language arts/literacy and math in an efficient, valid, and fair way
- The implementation of a broad-based test fairness agenda that continues College Board's practice of ensuring that the SAT Suite is a valid and fair assessment of all students' skills and knowledge
- The implementation of a wide-ranging test accessibility agenda that includes Bluebook's adherence to universal design principles, the provision of universal tools to all students during testing, and the availability of accommodations and supports for students who require them to access and respond to the test content
- Scores and score interpretation tools that provide clear, actionable information to students and their families, teachers, and other stakeholders
- Score reports that link students to both useful test data and a range of college and career opportunities and next steps
- Free, world-class practice opportunities that familiarize students with Bluebook, prepare them to answer test items successfully, and help them develop durable skills and knowledge needed for college and career readiness

The following subsections provide an overview of each of these central elements, most of which are discussed in greater depth throughout this manual.

## Digital Testing

The digital SAT Suite represents College Board's full shift to digitally based testing for its flagship college and career readiness assessments. All students—with the important exception of those requiring paper-based accommodations for fair access to the tests—now take SAT Suite tests digitally. This embrace of digital testing for the SAT Suite offers several critical benefits to those who take the tests, administer the tests, and use the tests' data. Refer to Chapter 4, Test Administration and Security, for a fuller discussion of the features and benefits offered by the digital test.

College Board administers the digital SAT Suite on Bluebook, a customized digital testing application that is the same platform used to successfully to deliver the AP Exams digitally. Having a well-vetted exam app allows College Board to fully meet SAT Suite users' needs and to respond in an agile manner by quickly making updates and refinements as needed. Refer to Chapter 4, Test Administration and Security, for an in-depth discussion of Bluebook and the ease with which it allows students to take the assessments.

## Multistage Adaptive Testing

For the digital SAT Suite, College Board has shifted from a linear testing model as the primary mode of administration to an adaptive one. In linear testing—the traditional approach for the SAT Suite—a student is given a test form with an array of items that has been set prior to test day and does not change based on the student's performance during the test. In an adaptive test model, the test delivery application instead adjusts the difficulty of the items given to students during the test based on the performance of individual test takers. These adjustments help ensure that any given student on test day is being administered items of difficulty levels appropriate to their level of achievement.

College Board employs a multistage adaptive testing model for the selected-response (Reading and Writing and Math) sections of the digital-suite tests. In this model, test content in each section is organized into two *stages*, each composed of a *module* of test items comprising half of the section's items. After answering an initial (routing) module of items representing a broad mix of easy, medium-difficulty, and hard items, students are routed to either a lower- or higher-difficulty second-stage module based on their performance on the items in the initial module. The mechanisms of multistage adaptive testing as it applies to the digital SAT Suite are explained in greater depth in Chapter 4, Test Administration and Security.

## Embedded Pretesting

The digital SAT Suite incorporates embedded pretesting into its design. In embedded pretesting, a small number of pretest (unscored) items are included, or *embedded*, among the operational (scored) items. The inclusion of these items allows College Board to evaluate them for potential (operational) use in future administrations of the tests. Although they are not administered for a score, these pretest items are otherwise indistinguishable to students from the operational items on which their scores are based. This ensures that students give maximum attention and effort to these items, which enhances the predictive power of the pretest statistics yielded.

## Discrete Items

All questions on the digital SAT Suite are in a discrete (standalone) format, meaning that students are able to answer each question independently, without reference to a common stimulus such as an extended passage. This represents a departure from the paper and pencil SAT Suite, which used a combination of discrete questions and question sets. The use of discrete questions is discussed in more detail in Chapter 4, Test Administration and Security.

## Fairness

College Board is strongly committed to the indivisibility of the concepts of test validity (i.e., that a test is measuring what it is intended to measure) and test fairness (i.e., that a test affords an equal opportunity to all test takers to perform up to their level of achievement without hindrance). To put the matter simply, a test must be fair to be valid. As with the paper-based suite, test fairness considerations permeate the design, development, and administration of the digital SAT Suite. Comprehensive discussions of fairness and validity can be found in Chapter 7, Fairness, and Chapter 8, Validity, respectively.

## Accessibility

Accessibility is a critical aspect of test fairness. The digital SAT Suite advances the goal of maximal accessibility for all students through a wide range of measures, including the application of universal design principles, the provision of universal tools, and the availability of accommodations and supports for those students who require them. A fuller discussion of these issues appears in Chapter 4, Test Administration and Security, and Chapter 7, Fairness.

## Scores, Score Interpretation Tools, and Student Score Reports

The digital SAT Suite tests yield three scores[2]—a total score and two section scores—accompanied by test interpretation tools that allow test takers and their families, educators, and other stakeholders to make informed, data-based decisions about students' educational futures. Scores for all the assessments are on the same vertical scale, allowing for meaningful interpretations about students' academic growth as they move between testing programs within the suite. Student score reports not only provide easy access to performance information and interpretation aids but also facilitate connections to educational opportunities, such as information and resources about local two-year colleges, workforce training programs, and career options. See Chapter 5, Test Scoring and Reporting, for more information.

2 Students taking the digital SAT Essay as part of select U.S. school day administrations receive three additional Essay scores: Reading, Analysis, and Writing, each on a 2–8 scale. These scores are not combined with each other and do not contribute to the test's section or total scores.

## Practice

Productive practice for the digital SAT Suite is supported by a number of high-quality College Board–created or –supported resources, available to all students, most of them at no cost. To be productive, practice must familiarize students with the test itself, its response formats, and its delivery method as well as help students build on what they are already good at and address weaknesses where they exist. Bluebook onboarding, full-length and item-level practice, and skill/knowledge building support are designed to facilitate students' readiness for test day and to meet College Board's professional and ethical obligation to level the playing field so that all test takers have an equal chance to demonstrate their achievement on the digital SAT Suite.

College Board conceptualizes *practice* for the digital SAT Suite as operating at three main levels:

1. Digital assessment readiness, which is intended to make students familiar and comfortable with Bluebook and the manner in which answer responses are entered
2. Test wisdom, which is intended to acquaint students with the types of items they will encounter on the tests, determine whether they can or cannot answer such items correctly, and offer insights into ways students can improve their future test performance
3. Skill and knowledge building, which is intended to help students gain durable academic abilities useful for college, career, and life

Conceiving of practice in these ways serves students far more effectively than do traditional forms of "test prep" focused only on the middle layer in the above scheme. To be clear, providing all students with practice test items is a critically important element of ensuring fairness and equity in testing, but overfocusing on repetitive test- or item-level practice risks narrowing students' attention and the secondary curriculum itself to only those skills and knowledge elements directly measured by an assessment and to the ways and manner these elements are sampled on a given test. In a real sense, practice focused mainly on such test preparation runs the risk of conflating a proxy of the desired skills and knowledge—that is, performance on an assessment—with the goal of developing students' durable skills and knowledge through a rich, diverse educational experience.

Table 1.1 provides a synoptic look at the several layers of practice opportunities available for the digital SAT Suite, each of which is discussed in more detail below.

**TABLE 1.1** DIGITAL SAT SUITE PRACTICE OPPORTUNITIES

| Form of Practice | Focus | Digital SAT Suite Implementation |
|---|---|---|
| Digital assessment readiness | Prepare students to test on Bluebook | ▪ Digital Test Preview |
| Test wisdom | Prepare students to answer test items productively and offer insight into students' academic strengths and weaknesses | ▪ Sample test items (with answer explanations)<br>▪ Official full-length practice test forms<br>▪ Official Digital SAT Prep on Khan Academy<br>▪ *The Official Digital SAT Study Guide*™ (print book)<br>▪ SAT Suite question banks<br>▪ Score reports<br>▪ Skills Insight™ score interpretation<br>▪ Test implementation guide |
| Skill and knowledge building | Build durable skills and knowledge needed for college, career, and life | ▪ Official Digital SAT Prep on Khan Academy<br>▪ Classroom practice guides<br>▪ High-quality instructional materials |

**Digital assessment readiness.** Students preparing for one of the digital SAT Suite tests have access to Digital Test Preview, which acquaints them with the central features of Bluebook and the assessments and presents them with a small number of sample Reading and Writing and Math items. These sample items serve primarily to familiarize test takers with the kinds of items they will be administered on test day and how to properly enter their answers rather than assess students' readiness to answer such items successfully.

**Test wisdom.** Students taking one of the digital SAT Suite tests have ready access to a wide range of high-quality test wisdom resources, all of them provided at no cost with the exception of the print book. Table 1.2 provides an overview of these resources.

**TABLE 1.2** DIGITAL SAT SUITE TEST WISDOM RESOURCES

| Test Wisdom Resource | Description |
|---|---|
| Sample test items (with answer explanations) | These items (**satsuite.collegeboard.org/media/pdf/digital-sat-sample-questions.pdf**) serve to illustrate the range of skills and knowledge sampled on the digital SAT (and other tests in the suite) as well as the response formats used (multiple-choice and, for select Math items, student-produced response). |
| Official full-length practice forms | Digital adaptive test forms are available through Bluebook, College Board's exam app, allowing students to practice using the same interface and format most of them will use on test day. Linear (nonadaptive) test forms, with directions for determining scores, are also available in Bluebook or from College Board as downloadable PDFs. The PDF versions of practice test forms are recommended only for students who will test with paper-based accommodations on test day. Because these forms are nonadaptive, they must be somewhat longer to achieve the same level of measurement precision as their digital adaptive counterparts. Students can visit **satsuite.collegeboard.org/digital/digital-practice-preparation/practice-tests** to get started. |
| Official Digital SAT Prep on Khan Academy | Khan Academy (**khanacademy.org/digital-sat**) offers students the opportunity to practice on sequences of test items and receive feedback, including answer explanations. |
| *The Official Digital SAT Study Guide* (print book) | *The Official Digital SAT Study Guide* offers authoritative insights and advice regarding taking the digital SAT (information that applies generally across all the suite's exams) as well as paper-based test forms with which students can practice (although, as noted earlier, practice in Bluebook is recommended for most test takers). |
| SAT Suite question banks | These free digital resources allow users to search through a repository of released digital SAT Suite test items. The banks' contents are filterable along many dimensions, making it easy for users to find exactly the items they want. Both educator- and student-facing versions are available. Students can use their bank to select and download items for practice, test familiarization, and item-level review, while educators can use theirs for those purposes as well or for instructional planning and quiz and formative assessment development. The educator bank can be found at **satsuitequestionbank.collegeboard.org**. The student bank is available as part of My Practice, which can be reached via Bluebook or directly (**https://mypractice.collegeboard.org/login**). |
| Score reports | Score reports provide students with their scores, information about what their scores mean, and suggestions for next steps, such as additional practice and links to college and workforce training opportunities. |

| Test Wisdom Resource | Description |
|---|---|
| Skills Insight score interpretation | Skills Insight verbally describes the skills and knowledge in reading and writing and in math that test takers scoring in particular ranges are likely to know and to be able to demonstrate. The descriptions at each score band are empirically derived from an analysis of student performance on digital SAT Suite test items. Exemplar items by test section and score range help concretize the verbal descriptors. An overview of Skills Insight is available at **satsuite.org/digital-skills-insight**. |
| Test implementation guide | This resource (**satsuite.org/digital-teacher-implementation-guide**), developed primarily for teachers, details the design of the digital SAT Suite and offers suggestions to educators looking to incorporate test preparation as part of their classroom instruction. |

**Skill and knowledge building.** College Board, in partnership with Khan Academy and others, makes a range of skill- and knowledge-building resources available for free. These resources are aimed at developing students' durable knowledge and skills rather than directly at preparing students for test day. Table 1.3 provides an overview of these resources.

**TABLE 1.3** DIGITAL SAT SUITE SKILL- AND KNOWLEDGE-BUILDING RESOURCES

| Skill- and Knowledge-Building Resource | Description |
|---|---|
| Official Digital SAT Prep on Khan Academy | In addition to providing test preparation activities, Khan Academy (**khanacademy.org/digital-sat**) offers students a range of high-quality skill- and knowledge-building activities, including numerous videos and articles that target specific areas where students might need additional support. |
| Classroom practice guides | These guides, designed primarily for teachers, are collections of essays on select topics written by experts in English language arts/literacy (**satsuite.org/digital-classroom-practice-English**) and math (**satsuite.org/digital-classroom-practice-math**). The essays discuss critical college and career readiness requirements and how instruction can be designed to support all secondary students obtaining those competencies. Sidebars draw links between the essays' topics and how those topics are addressed on the digital SAT Suite tests. |
| High-quality instructional materials | College Board offers a wide range of high-quality instructional materials through its Pre-AP® and AP Programs. These programs support all students' attainment of critical college and career readiness prerequisites. |

## 1.3.2 Benefits of the Digital SAT Suite

The digital SAT Suite is part and parcel of College Board's larger mission to promote access and opportunity. The digital SAT Suite is built on the firm foundations of the paper and pencil SAT Suite to make the digital-suite exams:

- Easier to take
- Easier to give
- More secure
- More relevant

Each of these guiding principles is discussed in turn in the following subsections.

### Easier to Take

In a number of important ways, the digital SAT Suite tests are easier to take than their paper and pencil predecessors. The digital tests themselves are roughly an hour shorter, and pre- and posttest activities and administrative time have been significantly reduced, meaning that test day is a much more streamlined experience for all involved. Students can take the digital tests on a wide range of devices, including personal or school-managed Windows laptops or tablets, Mac laptops, iPads, and school-managed Chromebooks. Because not all students will have ready access to a digital device on which to test, College Board is committed to lending a device to any student testing on the weekend who needs one. For more information, please visit **satsuite.collegeboard.org/digital/device-lending**.

Digital-suite test items, while preserving the rigor of the paper and pencil SAT Suite tests, are concise and focused, facilitating their delivery on digital devices. Bluebook renders these items and the tests themselves in a fluid, intuitive way based on principles of universal design; features numerous tools, such as the built-in Desmos Graphing Calculator as well as the ability to annotate and to flag questions, that all students may opt to use; and makes available a wide range of accommodations and supports for those students who require them to access the tests and their content.

Data College Board collected from participants throughout 2023, the first year of operational digital administrations, strongly support the claim that the digital SAT Suite tests are easier to take:

- 76% of surveyed test takers (n=107,385) reported an excellent or good experience taking a digital SAT Suite test.
- 84% of students who had previously taken a paper-based SAT (n=4,957) reported a better test-taking experience with the digital SAT.
- 80% of students who had previously taken a paper-based SAT (n=888) stated they felt less overwhelmed by the digital SAT.
- 89% of students who had taken other digital assessments (n=62,239) reported a better experience (48%) or the same quality of experience (41%) taking the digital SAT.
- 95% of students (n=120,790) felt Bluebook was easy to use.

## Easier to Give

In their digital form, the SAT Suite assessments are also easier to give than ever before. Gone are the days of shipping, securing, unpacking, distributing, collecting, and repacking test materials, all of which carried with them attendant operational and security risks. The tests themselves have fewer separately timed sections, thereby easing administration, and exam timing is handled by Bluebook itself, not the proctor. The Test Day Toolkit app created by College Board makes the remaining test administration tasks much easier for proctors and test center coordinators as well. Having significantly shorter tests means it is easier for schools administering digital SAT Suite tests as part of the school day to give those exams on their schedule, not College Board's, and the various innovations College Board has introduced by the move to digital mean that more test administrations and wider, more flexible school day testing windows can be introduced, furthering the goal of test access.

With the digital suite, College Board has also taken seriously the concern that not all schools or other test centers have sufficient, or sufficiently reliable, internet access to support large numbers of students simultaneously and continuously accessing Wi-Fi or other networks during testing. And, of course, device batteries can fully drain at inconvenient times. That is why College Board designed Bluebook to be tolerant of momentary interruptions in connectivity (whether network or battery related) without losing students' work or time. Should students experience brief interruptions in their connectivity, they can quickly resume with no loss of testing time; should their device battery fully drain during testing, they can simply plug in, restart, and resume testing without loss of either testing time or their work, as Bluebook automatically saves their responses.

Data College Board collected from participants throughout 2023 back up the claim that the digital-suite tests are easier to give:

- 75% of test center coordinators (n=13,461) and 80% of staff (n=39,944) rated their experience with the digital SAT as good or excellent.
- 87% of test center coordinators and staff (n=46,166) believe students experienced minimal distraction while taking the digital SAT.
- 91% of test center coordinators and staff (n=46,162) felt their test center was able to provide assistance to students who needed it.
- 92% of test center coordinators and staff (n=46,208) think students understood the expectations of a digital SAT.
- 93% of test center coordinators and staff (n=45,605) stated they felt very or somewhat prepared to administer the digital SAT.

## More Secure

The tests of the digital SAT Suite are also more secure than the paper and pencil tests they have replaced. As mentioned above, the switch to digital has eliminated the paper handling that not only places burdens on test administrators but also creates security risks. Bluebook also displays only one test item at a time, making it much more difficult for bad actors to surreptitiously photograph or otherwise copy test content. Most critically, though, the digital SAT Suite assessments have been designed and developed such that each student is administered a highly comparable but unique version of the test. This innovation greatly diminishes any value in students copying from their test-taking neighbors or scouring the internet for leaked test forms.

## More Relevant

The digital SAT Suite tests are also more relevant for all students than ever before. College Board has always sought to reflect in its test materials the widest possible range of information, ideas, and perspectives, and, to a large extent, the paper and pencil versions of the SAT Suite achieved those goals. However, the use on the paper-based suite of a relatively few extended (multiparagraph) passages as the basis for many test items placed a hard limit on the range of texts that could be presented.

With the digital tests, the number and variety of contexts serving as the basis for test items have been greatly increased. This means that there are many more opportunities for the tests to represent the diversity of people, experiences, and interests in the United States and around the world. This, in turn, greatly increases the chances that students on test day will encounter passages that they find meaningful and personally interesting, as College Board's early research on student perceptions of the digital tests has suggested. College Board believes the end result will be more engaged test takers whose scores reflect their best efforts.

Data College Board collected from participants throughout 2023 support the claim that the digital-suite tests are more relevant.

- 85% of surveyed students (n=868) said the Reading and Writing passages were less stressful than their paper-based counterparts.
- 80% of surveyed students (n=853) said the digital passages were more relatable than paper-based SAT passages.
- 80% of surveyed students (n=887) said the digital passages were more engaging than the passages found in the paper-based SAT.

Chapter 2

# Test Specifications

## 2.0 Introduction

Test specifications define the key parameters of a given assessment. Such parameters include the purpose or purposes of the assessment; what concept or concepts the assessment is designed to measure (its *construct* or *constructs*); what assertions about test taker performance the assessment is designed to collect evidence in support of (its *claims*); what skills, knowledge, and abilities the assessment is intended to measure attainment of (its *content specifications*); and what psychometric characteristics its questions, problems, and tasks are required to have (its *statistical specifications*). The twin goals of test specifications are, first, to lay out in precise detail what the assessment is designed to measure and how this is to be done so that the validity of its design can be fairly evaluated and, second, to provide the replicable specificity needed to ensure that test-level events (*forms*) of the assessment are as closely parallel as possible, thus ensuring a high degree of standardization. A presentation of a test's specifications is a crucial element of the validity argument undergirding that test, as the specifications establish the *what* and *how* of a test in conformity with the test's purpose, constructs, and claims, thereby establishing the manner by which a test measures what it is intended to measure.

This chapter provides an overview and discussion of the test specifications for the digital SAT Suite assessments. The chapter is divided into three major sections. Section 2.1, Key Concepts, provides background information on topics important to subsequent discussion, including the intended purposes and uses of the digital SAT Suite and the concepts of *constructs* and *claims* as they apply to assessment. Section 2.2, Test Section Overviews, discusses central features of the various section-level components of the suite's tests. Section 2.3, Statistical Specifications, address the substantive and measurement properties of the tests in the suite.

## 2.1 Key Concepts

Three important concepts included in the chapter's subsequent discussion are worth defining at the outset: the *purpose(s)* and *intended use(s)* of a test, the *construct(s)* intended to be measured by the test, and the *claim(s)* about test taker performance the assessments are designed to collect evidence in support of. Each concept is discussed first in general terms and then with respect to the digital SAT Suite.

### 2.1.1 Purposes and Intended Uses and Interpretations

A test's *purpose* or *purposes* define the goal(s) of the assessment and indicate its intended uses as well as uses for which the assessment was not intended. Clear and precise definition of a test's (or testing suite's) purpose(s) and intended use(s) is critical because such define the broad parameters of the assessment, serve as the foundation for an argument about the assessment's validity, and allow potential users to fairly evaluate whether the assessment is suitable for their needs.

The primary purpose of the tests of the digital SAT Suite is to determine the degree to which students are prepared to succeed both in college and careers. All assessment content, which has been developed based on high-quality research identifying the knowledge and skills most essential to college and career readiness and success, aligns with this core purpose. Each test within the digital SAT Suite is designed to collect evidence from student performance in support of broad *claims* (defined more precisely below) about what students know and can do, and each claim is aligned to the primary purpose of assessing college and career training program readiness. The resulting scores provide meaningful information about a student's likelihood of succeeding in college—information that, used in conjunction with other data (such as high school grades) and in the context of where a student lives and learns, can contribute to decisions about higher education admission and placement.

Although the core purpose of the digital SAT Suite is college and career readiness assessment, the suite's data are employed for many purposes by a range of users, notably higher education and K–12 educators, and students. In keeping with best practices and professional standards (AERA et al., 2014), the digital SAT Suite's intended uses and interpretations are discussed in the following paragraphs, with a rationale presented for each use.

**Evaluating and monitoring students' college and career readiness** (for use by K–12 educators and students). The digital SAT's empirically derived College and Career Readiness Benchmarks ("SAT benchmarks") serve as challenging, meaningful, and actionable indicators of students' college and career readiness. States, districts, and schools use the SAT benchmarks to monitor and determine what proportion of their student body has a high likelihood of success in 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 established for the digital PSAT-related assessments 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. For more details on the benchmarks, see Chapter 5, Test Scoring and Reporting. Additionally, the PSAT 8/9 and the SAT are used to satisfy federal accountability requirements in 8th and 11th grade in several states.

**Monitoring student progress through a vertically scaled suite of assessments** (for use by K–12 educators and students). Every test in the digital SAT Suite is reported on the same vertical scale, with the digital SAT as the capstone measure. The scales for the digital SAT have been established through concordance studies using samples of U.S.- and non-U.S.-based students, and the scales for the digital PSAT/NMSQT, PSAT 10, and PSAT 8/9 tests have been linked to the digital SAT scale through vertical scaling studies using U.S.-based students. Establishing the scales in this manner allows for appropriate inferences regarding a student's academic growth and their progress toward college and career readiness from year to year prior to them taking the digital SAT. One is then able to make statements about a student's level of preparedness for college and careers based on digital SAT performance. Students can track their own progress by using information provided in their score report to identify instructional areas needing improvement and then engage in practice and learning opportunities that will help them become more prepared for college-level work. For details on scores, score interpretation, and student score reports, see Chapter 5, Test Scoring and Reporting; for details on practice opportunities, see Chapter 1, Overview of the Digital SAT Suite.

**Contributing to high school course placement decisions** (for use by K–12 educators and students). All assessments across the digital SAT Suite provide information about a student's readiness for particular AP courses. AP Potential is a free, online tool that allows schools to generate rosters of students who are likely to score a 3 or higher on a given AP Exam based on their performance on the PSAT 8/9, PSAT/NMSQT, PSAT 10, or SAT. SAT Suite scores are stronger predictors of students' AP Exam scores than are more traditional factors such as high school grades and grades in same-discipline coursework. AP Potential should never be used to discourage a motivated student from registering for an AP course. The AP Program encourages schools to use a variety of factors, including grades, student motivation, and teacher recommendations, when registering students for AP courses.

**Connecting students to career possibilities** (for use by students). Discovering career options is a driving force as students make decisions about their future. Every career requires a set of skills, the attainment of which can be measured. College Board has worked with experts in occupations and labor market data to map the reading, writing, and math skills and knowledge measured on the SAT and the PSAT-related assessments to the literacy and numeracy requirements of a thousand different careers. To help all students consider the full range of vocational options open to them, digital SAT Suite score reports include the Career Insights Snapshot, which lists careers in a student's state that are connected to the student's assessment performance. Each listed career has a bright outlook, pays a living wage in the state, and requires some form and level of postsecondary education. These careers are presented as examples and are neither formal recommendations nor the only career options that students should consider.

**Making college admission and college course placement decisions** (for use by higher education). In conjunction with other sources of data, such as high school grades, as well as contextual information about where students live and learn, as provided through College Board's Landscape® tool (**professionals.collegeboard.org/landscape**), the digital SAT is intended for use in college admission and course placement decisions. The digital SAT provides information on a student's level of preparedness for college-level work, which helps admission professionals make more informed selection decisions. Once students are admitted, SAT results offer useful and valid ways to support students on campus, including informing course placement and academic major decisions and helping identify students in need of academic support.

**Contributing to scholarships and other awards** (for use by higher education and nonprofit organizations). Scores from the digital SAT Suite often inform decisions that colleges and nonprofit programs make in relation to academic awards, scholarships, and other aid.

**Connecting students to college and scholarship opportunities** (for use by students). As students consider college options and explore scholarship opportunities, they may choose to use their digital SAT score band in a variety of ways:

Students can explore colleges and scholarships that may be a good match for them by searching more than 4,400 institutions and 6,000 scholarship programs using an SAT score range filter on BigFuture (**bigfuture.collegeboard.org**), College Board's free online planning guide through which students can explore careers, plan for college, and explore methods to pay for college.

Students who take in-school assessments may choose to hear from nonprofit accredited colleges and universities (domestic and international), nonprofit scholarship providers, and government agencies administering educational programs that might be a good match for them through Connections.[3] College Board will deliver messages from these organizations to students via the BigFuture School app (**bigfutureschool.org**), email, and postal mail when available (if the student chooses to download BigFuture School, and/or provide their email and/or postal address when they take the eligible test[4] or later in BigFuture School mobile app, all of which are optional). No personally identifiable information is shared unless and until the student chooses to share it directly with an organization.

All students have the option to join Student Search Service™ so that interested colleges and scholarship programs can reach out to them directly. When students join Student Search Service, their digital PSAT/NMSQT, PSAT 10, and/or SAT score bands, AP Exam score bands, if any, along with other information about them and their interests, are used and shared with participating nonprofit organizations to help these organizations looking for students who are a good match for them. These organizations may send students (and their parent/guardian if they have opted in) email and postal mail informing them about their educational, financial aid, and scholarship opportunities.

**Helping underrepresented students be seen by colleges**. The College Board National Recognition Program awards academic honors to high-performing underrepresented students. The five national recognition programs include the National First-Generation Recognition Program, the National African American Recognition Program, the National Hispanic Recognition Program, the National Indigenous Recognition Program, and the National Rural and Small Town Recognition Program. Students who take eligible administrations of the digital PSAT/NMSQT, digital PSAT 10, or AP Exams and meet the score requirements are considered for the awards, which are a tangible way to help students be seen by colleges and support colleges' recruitment strategies.

3 A school, district, or state may choose not to offer Connections to its students.

4 For more, see **privacy.collegeboard.org/program-specific-privacy-policies/bigfuture-connections/connections.**

Implicit in the preceding discussion are limits on the intended uses of the digital-suite tests. The suite is intended to open doors for students and to help them gain access to opportunities that they have earned through their hard work. It is therefore inappropriate to use digital SAT Suite scores as a veto on students' educational or vocational aspirations. When interpreted properly, data from tests such as those of the digital SAT Suite can make valuable contributions to helping students meet their academic and career goals, but test scores should never be the sole basis for highly consequential decisions about students' futures. Digital SAT Suite scores, therefore, should be considered alongside other factors, including high school grades and where students live and learn, when evaluating students' achievement or potential.

Digital SAT Suite scores should also not be used as the single measure to rank or rate teachers, educational institutions, districts, or states. Users should exercise care when attempting to interpret test results for a purpose other than the intended purposes described above. College Board is not aware of any compelling validation evidence to support the use of any of the digital SAT Suite assessments, or other educational achievement measures, as the principal source of evidence for teacher or school leader evaluation. Assessment data, when 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. For further examples of uses of College Board test scores that should be avoided, see *Guidelines on the Uses of College Board Test Scores and Related Data* (College Board, 2018).

### 2.1.2 Constructs

A crucial foundation of assessment validity is a clear definition of a test's construct. A *construct*, in this sense, is "the concept or characteristic that a test is designed to measure" (AERA et al., 2014, p. 11). Clearly defining the construct—in the case of the digital SAT Suite, at the test section level—serves two main purposes. First, such a definition establishes the basis for and guides the development of test specifications, both content and statistical, which, in turn, set the parameters for what will be measured and how by the assessment. Second, the construct definition helps identify what should not be measured by the assessment. Such elements are considered *construct irrelevant*—that is, they should not have an impact on test performance, and any effects they might have should be minimized when they cannot be eliminated. In the case of the digital SAT Suite, such construct-irrelevant factors include, but are not necessarily limited to, students' race/ethnicity, gender and sexual identities, home region, home culture, family income, and physical ability.

It should be clear, then, that construct definition is not merely a technical matter but is, in fact, a cornerstone of both test validity and test fairness. By defining what should and should not be measured, developers can help ensure that a given test measures what it is intended to measure and minimize the possibility that other factors not intended to be assessed have impacts on performance. Indeed, "the central idea of fairness in testing is to identify and remove construct-irrelevant barriers to maximal performance for any examinee," and "removing these barriers allows for the comparable and valid interpretation of test scores for all examinees" (AERA et al., 2014, p. 63).

Section 2.2, Test Section Overviews, specifies the constructs for the components of the digital SAT Suite assessments.

### 2.1.3 Claims

Like constructs, *claims* serve to establish parameters for what is tested and how. "In assessment," Mislevy and Riconsente (2005) write, "we want to make some claim about student knowledge, skills, or abilities (KSAs), and we want our claims to be valid" (p. 1). The purposes of establishing claims, then, are first to precisely identify what students should be able to demonstrate to be considered successful on the construct of interest and, second, to build a test (or test section) to collect evidence in support of those claims.

Section 2.2, Test Section Overviews, specifies the claims associated with the components of the digital SAT Suite tests.

## 2.2 Test Section Overviews

All assessments of the digital SAT Suite—the SAT, PSAT/NMSQT, PSAT 10, and PSAT 8/9—consist of two sections: a Reading and Writing section and a Math section. The digital SAT (only) may also be accompanied by the digital SAT Essay, a direct-writing assessment administered as part of select U.S. school day administrations at the request of particular state education departments. This subsection provides an overview of each of these components, with additional details to follow in the chapter.

### 2.2.1 The Reading and Writing Section

The Reading and Writing section of the digital SAT Suite of Assessments is designed to measure students' attainment of critical college and career readiness prerequisites in literacy in English language arts as well as in various academic disciplines, including literature, history/social studies, the humanities, and science. The Reading and Writing section focuses on key elements of comprehension, rhetoric, and language use that the best available evidence identifies as necessary for postsecondary readiness and success. Over the course of a Reading and Writing section of one of the digital SAT Suite assessments, students answer multiple-choice questions requiring them to read, comprehend, and use information and ideas in texts; analyze the craft and structure of texts; revise texts to improve the rhetorical expression of ideas; and edit texts to conform to core conventions of Standard English.

#### Construct

The construct (concept) intended to be measured by the digital SAT Suite's Reading and Writing section is literacy achievement relative to core college and career readiness requirements in English language arts as well as in the academic disciplines of literature, history/social studies, the humanities, and science.

## Claims

The Reading and Writing section is designed to elicit evidence from student test performance in support of four broad claims about students' literacy achievement. To be successful on the Reading and Writing section, students must be able to:

- Demonstrate understanding of **information and ideas** in texts across a range of academic disciplines and complexities aligned with college and career readiness requirements
- Effectively evaluate the **craft and structure** of texts, including demonstrating understanding and proficient use of high-utility academic vocabulary in context
- Revise the **expression of ideas** in texts to enhance communicative power in accordance with specified rhetorical goals
- Edit texts in accordance with **Standard English conventions** in order to meet academic and workplace expectations regarding the use of standardized expression

## Content Domain Structure

In close correspondence with the claims listed above, items on the Reading and Writing section represent one of four content domains:

- **Information and Ideas**, for which students must use comprehension, analysis, and reasoning skills and knowledge as well as what is stated and implied in texts (including in any accompanying informational graphics) to locate, interpret, evaluate, and integrate information and ideas
- **Craft and Structure**, for which students must use comprehension, vocabulary, analysis, synthesis, and reasoning skills and knowledge to use and determine the meaning of high-utility academic words and phrases in context, evaluate texts rhetorically, and make supportable connections between multiple topically related texts
- **Expression of Ideas**, for which students must use revision skills and knowledge to improve the effectiveness of written expression in accordance with specified rhetorical goals
- **Standard English Conventions**, for which students must use editing skills and knowledge to make text conform to core conventions of Standard English sentence structure, usage, and punctuation

Each Reading and Writing test item is classified as belonging to a single content domain. The first two content domains—Information and Ideas and Craft and Structure—primarily address reading-related skills and knowledge, while the second two content domains—Expression of Ideas and Standard English Conventions—primarily address writing-related skills and knowledge. This division into reading- and writing-focused content domains is purely heuristic and has no substantive bearing on the test structure or the scores reported. The divisions into "reading" and "writing" are, in fact, somewhat porous, and the test section itself may fairly be thought of as a blended reading and writing assessment. Within each domain, items address a number of skill/knowledge testing points; as discussed in Appendix A, items from all four content domains appear in each test module.

## Key Features

Table 2.1 summarizes many of the key features of the Reading and Writing section. These features are discussed following the table.

**TABLE 2.1 DIGITAL SAT SUITE READING AND WRITING SECTION: HIGH-LEVEL SPECIFICATIONS**

| Characteristic | Digital SAT Suite Reading and Writing Section |
|---|---|
| Administration | Two-stage multistage test design; one Reading and Writing section administered via two separately timed modules |
| Test length (number of operational and pretest items) | First module: 25 operational items and 2 pretest items<br>Second module: 25 operational items and 2 pretest items<br>Total: 54 items |
| Time per module | First module: 32 minutes<br>Second module: 32 minutes<br>Total: 64 minutes |
| Average time per item | 1.19 minutes |
| Score reported | Section score (constitutes half of total score)<br>SAT: 200–800, in 10-point intervals<br>PSAT/NMSQT and PSAT 10: 160–760, in 10-point intervals<br>PSAT 8/9: 120–720, in 10-point intervals |
| Item format used | Discrete; four-option multiple-choice with a single best answer (*key*) |
| Passage subject areas | Literature, history/social studies, the humanities, science |
| Word count | 25–150 (six-character) words per stimulus text (or pair of texts) |
| Informational graphics | Tables, bar graphs, line graphs |
| Text complexity bands | Grades 6–8, grades 9–11, grades 12–14 (grades 12–14 excluded from PSAT 8/9) |

The following discusses the key features identified in Table 2.1.

### Administration

The digital SAT Suite Reading and Writing section is administered using a straightforward multistage adaptive test (MST) design consisting of two equal-length and separately timed portions, or *stages*. Each stage consists of a *module* of test items. The first (*routing*) module in each test form consists of a broad mix of easy, medium-difficulty, and hard items. The second module also contains a mix of easy, medium-difficulty, and hard items, but the average item difficulty is targeted to the test taker's performance on the items in the first module. Specifically, a test taker receives a second module that consists of items that, on average, are either of higher or lower difficulty than those in the first module. Student performance on all operational (i.e., non-pretest) items across both modules is used to calculate the section score.

Items from all four of the section's content domains appear in each test module in the sequence depicted in Table 2.2.

**TABLE 2.2 DIGITAL SAT SUITE READING AND WRITING SECTION: ITEM SEQUENCE**

| Reading and Writing Module | Content Domain Sequence |
|---|---|
| Module 1 | Craft and Structure items |
| | Information and Ideas items |
| | Standard English Conventions items |
| | Expression of Ideas items |
| Module 2 | Craft and Structure items |
| | Information and Ideas items |
| | Standard English Conventions items |
| | Expression of Ideas items |

The skill/knowledge elements assessed by the section's content domains are discussed in detail in Appendix A.2.3.1. The Reading Section

Within each content domain except Standard English Conventions, items are ordered first by skill/knowledge element and then by item difficulty from easiest to hardest. Standard English Conventions items are ordered from easiest to hardest irrespective of skill/knowledge element tested. Placing items testing similar skills and knowledge together reduces task switching and makes it easier for students to budget their time, while the ordering of items from easiest to hardest (within the placement restriction noted above) means that students can build confidence as they move from relatively straightforward to more demanding assessment tasks. Embedded pretest items appear alongside items testing the same skill/knowledge element (or, in the case of Standard English Conventions pretest items, among the operational items in that content domain).

Students may navigate freely through a given module of items; this allows them to preview upcoming items within the module or flag others to return to should time permit. However, once Bluebook has moved them to the second module in the Reading and Writing section, students may not return to the items in the first module, nor may students return to the Reading and Writing section once Bluebook has moved them to the Math section.

### Test Length by Number of Items

The Reading and Writing section consists of 54 items, which are divided into two equal-length modules, one for each of the section's two stages. Of each module's 27 items, 25 are *operational*, meaning that students' performance on them contributes to their section score, and 2 items are *pretest*. Answers to pretest items do not contribute in any way to students' Reading and Writing section score; rather, their purpose is to collect student performance data that will be used by College Board to help assess whether these items are suitable for operational use at a future time. To students, pretest items appear the same as operational items, thereby ensuring students' maximal effort on the former. The number of pretest items embedded in the Reading and Writing section is minimized to limit the impact of answering nonoperational items on students' test taking.

### Time Per Section and Module

Students have a total of 64 minutes to complete the Reading and Writing section. This time is divided equally between the two modules, meaning that students have 32 minutes to answer the items in a given module. Once time has expired for the first module, students are automatically advanced to the appropriate (higher- or lower-difficulty) second module and may not return to the items in the first module.

### Average Time Per Item

Students have, on average, 1.19 minutes to answer each Reading and Writing test item.

### Score Reported

Students receive a section score based on their overall performance on the Reading and Writing section. This score is scaled from 200–800 for the SAT, 160–760 for the PSAT/NMSQT and PSAT 10, and 120–720 for the PSAT 8/9, each in 10-point intervals. This section score is added to the Math section score to determine students' total score for the assessment.

### Item Format Used

All Reading and Writing test items are in the four-option multiple-choice format with a single best answer known as the *keyed response* or simply the *key*. These items are considered *discrete* because each has its own passage (or passage pair) serving as a stimulus and because no item is linked to any other item in the section.

### Passage Subject Areas

Passages in the Reading and Writing section, which serve as the basis for answering test items, represent the subject areas of literature, history/social studies, the humanities, and science. Topic-specific prior knowledge is not assessed.

### Word Count

The passage (or passage pair) accompanying each Reading and Writing test item ranges from 25 to 150 words. To ensure that this word count is not unduly influenced by the presence of many long or short words in a given passage (or passage pair), a *word* in this sense is considered six characters, with characters including letters, numbers, spaces, and symbols (including punctuation). A standardized word count is thus calculated by dividing the passage's total character count by six. When a single test item uses a pair of short passages as the stimulus, the total word count for both passages must fall within the specified range.

### Informational Graphics

Select passages in the Reading and Writing section are accompanied by an informational graphic. The goal of the inclusion of such graphics is to authentically assess students' ability to locate and interpret data and to use these data effectively to answer an associated item. For the Reading and Writing section, informational graphics are limited to tables, bar graphs, and line graphs, as these are the most common ways to display data in the subject areas sampled by the section. Items on the section associated with informational graphics do not require students to perform calculations (and calculators are not permitted on the section). Instead, students must use their quantitative and disciplinary literacy skills and knowledge to find relevant data in graphics, make reasonable interpretations of those data, and use those data along with information and ideas in associated passages to reach supportable conclusions.

### Text Complexity

The overall text complexity of Reading and Writing passages is aligned with college and career readiness requirements, with individual passages (or passage pairs) representing one of three complexity levels: grades 6–8, grades 9–11, or grades 12–14. Passages of the highest text complexity band are excluded from PSAT 8/9, as these texts are generally too challenging to contribute materially to an assessment of eighth and ninth graders' literacy achievement. Bands rather than grades are used to conceptualize text complexity for two main reasons. First, text complexity expectations at the secondary level are relatively compressed across grades, resulting in significant overlap between and among grade-specific expectations. Second, even minor changes to the wording of passages as short as those used in the Reading and Writing section can significantly influence the results yielded by quantitative text complexity measures, so the use of bands permits some necessary flexibility in passage development while maintaining an appropriate and easily understood range of text complexities across the section.

Reading and Writing passages' complexity is determined using both quantitative and qualitative measures. To ascertain a text's complexity, College Board uses both a robust quantitative measure and a qualitative rubric. The quantitative tool takes a text of any size and produces three measurements: Syntactic Complexity, Academic Vocabulary, and an overall model prediction. The Syntactic Complexity measure evaluates more than two dozen text attributes, including mean sentence length before the sentence root, the number of dependent clauses per sentence, and intersentence cohesion. The Academic Vocabulary measure evaluates more than a dozen text attributes, including the average frequency with which words in the text appear in a corpus of college-level textbooks, the average age at which people typically acquire the words in the text, and the average concreteness of words in the text. Syntactic Complexity and Academic Vocabulary are calculated values. The model prediction is inferred from a model that has been trained on the CommonLit dataset binned into the ranges used on the digital SAT Suite.

## Passages and Items

The passages serving as stimuli for Reading and Writing items represent a range of academic disciplines and text complexities aligned to college and career readiness requirements. Passages in the literature subject area are excerpted from previously published third-party sources for which permission to use has been obtained if necessary (i.e., not for works in the public domain), while all other Reading and Writing passages are written specifically for the tests in a way that maintains verisimilitude by including factual information and reflecting the norms and conventions (e.g., structure, language patterns, vocabulary, style and tone) of authentic texts in the disciplines. Associated questions pose tasks similar to those assigned in rigorous secondary school classes, entry-level, credit-bearing college courses, and workforce training programs. These tasks include comprehending information and ideas conveyed explicitly and implicitly in texts from various academic disciplines and of varying complexities; demonstrating command of textual and quantitative evidence; determining the meaning of and skillfully using high-utility academic (tier two) words and phrases in context; revising texts to accomplish particular rhetorical purposes; and editing texts so that they conform to core conventions of Standard English sentence structure, usage, and punctuation.

All Reading and Writing items are four-option multiple-choice in format, with a single best answer to each question. This best answer is known as the *keyed response* or the *key*, while the three alternatives are known as *distractors*. Each distractor represents a common error that students might reasonably make in answering the item, so distractors are intended to be plausible to degrees varying depending on the intended difficulty of the item; however, no distractor is meant to compete with the key as the best response for students with the targeted level of reading and writing achievement.

### Variations by Testing Program

Given that the digital SAT Suite assessments are intended to measure attainment of a fixed target—college and career readiness—and because evidence informing the broader literacy construct tends to suggest only relatively small variations in skill and knowledge requirements across the secondary grades, the Reading and Writing sections across all testing programs (SAT, PSAT/NMSQT and PSAT 10, and PSAT 8/9) are the same in terms of content specifications with one exception: PSAT 8/9 does not include passages from the highest (grades 12–14) text complexity band. Statistical specifications vary by testing program, meaning that the range of difficulty of items students are administered differs from program to program. This, along with the multistage adaptive nature of testing and the PSAT 8/9 text complexity constraint, ensures that students taking the various tests are given age- and attainment-appropriate items, with allowances for higher- and lower-achieving students within each test-taking population to demonstrate the full extent of their ability.

## 2.2.2 The Math Section

The Math section of the digital SAT Suite assessments is designed to measure students' attainment of critical college and career readiness prerequisites in math. The digital SAT Suite Math section focuses on key elements of algebra, advanced math, problem-solving and data analysis, and geometry and (SAT, PSAT/NMSQT, and PSAT 10 only) trigonometry that the best available evidence identifies as necessary for postsecondary readiness and success. Over the course of the Math section of one of the digital SAT Suite assessments, students answer multiple-choice and student-produced response (SPR) questions that measure their fluency with, understanding of, and ability to apply the math concepts, skills, and practices that are most essential for readiness for entry-level postsecondary work.

### Construct

The construct (concept) intended to be measured by the digital SAT Suite's Math section is math achievement relative to core college and career readiness requirements. Although literacy achievement is not directly measured, students are still required to employ such skills and knowledge to a limited, carefully constrained extent when solving math problems set in context.

## Claims

The Math section is designed to elicit evidence from student test performance in support of four broad claims about students' math achievement. To be successful on the Math section, students must be able to do the following:

Analyze, fluently solve, interpret, and create linear equations and inequalities as well as analyze and fluently solve systems of equations using multiple techniques (**Algebra**)

Demonstrate attainment of skills and knowledge central for successful progression to more advanced math courses, including analyzing, fluently solving, interpreting, and creating equations; including absolute value, quadratic, exponential, polynomial, rational, radical, and other nonlinear equations, as well as analyzing and fluently solving systems of linear and nonlinear equations in two variables (**Advanced Math**)

Apply quantitative reasoning about ratios, rates, and proportional relationships; understand and apply unit rate; and analyze and interpret one- and two-variable data (**Problem-Solving and Data Analysis**)

Solve problems that focus on perimeter, area, and volume; angles, triangles, and trigonometry; and circles (**Geometry and Trigonometry** [SAT, PSAT/NMSQT, and PSAT 10] / **Geometry** [PSAT 8/9])

These general suite-level claims are modified to some extent at the individual test program level, as demonstrated in Appendix A.

## Content Domain Structure

In close correspondence with the claims listed above and as appropriate for the age and attainment of the test-taking populations targeted by the various digital SAT Suite assessments, items on the Math section represent one of four content domains:

**Algebra**, for which students must analyze, fluently solve, and create linear equations and inequalities as well as analyze and fluently solve systems of equations using multiple techniques

**Advanced Math**, for which students must demonstrate attainment of skills and knowledge central for successful progression to more advanced math courses, including analyzing, fluently solving, interpreting, and creating equations, including absolute value, quadratic, exponential, polynomial, rational, radical, and other nonlinear equations, as well as analyzing and fluently solving systems of linear and nonlinear equations in two variables

**Problem-Solving and Data Analysis**, for which students must apply quantitative reasoning about ratios, rates, and proportional relationships; understand and apply unit rate; and analyze and interpret one- and two-variable data

**Geometry and Trigonometry** (SAT, PSAT/NMSQT, and PSAT 10) / **Geometry** (PSAT 8/9), for which students must solve problems that focus on perimeter, area, and volume; angles, triangles, and trigonometry; and circles

Some notable variations by testing program exist; see Appendix A.

Each item is classified as belonging to a single content domain. Within each domain, items address a number of skill/knowledge testing points, as demonstrated in Appendix A. Items from all four content domains appear in each module.

## Key Features

Table 2.3 summarizes many of the key features of the Math section. These features are discussed following the table.

**TABLE 2.3** DIGITAL SAT SUITE MATH SECTION: HIGH-LEVEL SPECIFICATIONS

| Characteristic | Digital SAT Suite Math Section |
|---|---|
| Administration | Two-stage multistage test design; one Math section administered via two separately timed modules |
| Test length (number of operational and pretest items) | First module: 20 operational items and 2 pretest items<br>Second module: 20 operational items and 2 pretest items<br>Total: 44 items |
| Time per module | First module: 35 minutes<br>Second module: 35 minutes<br>Total: 70 minutes |
| Average time per item | 1.59 minutes |
| Score reported | Section score (constitutes half of total score)<br>SAT: 200–800, in 10-point intervals<br>PSAT/NMSQT and PSAT 10: 160–760, in 10-point intervals<br>PSAT 8/9: 120–720, in 10-point intervals |
| Item formats used | Discrete; four-option multiple-choice with a single best answer (*key*) (≈75%) and student-produced response (SPR; ≈25%) |
| Context topics | Science, social studies, real-world topics |
| Word count | Approximately 30% of items in context; a majority of in-context items have 50 (six-character) words or fewer |
| Informational graphics | A wide range of data displays, geometric figures, and *xy*-plane graphs |
| Text complexity bands | N/A; see below |

The following discusses the key features identified in Table 2.3.

## Administration

The digital SAT Suite Math section is administered using a straightforward multistage adaptive test design consisting of two equal-length and separately timed portions, or *stages*. Each stage consists of a *module* of test questions. The first (*routing*) module in each test form consists of a broad mix of easy, medium-difficulty, and hard questions. The second module of questions also contains a mix of easy, medium-difficulty, and hard questions, but the average item difficulty is targeted to the test taker's performance on the items in the first module. Specifically, a test taker receives a second module that consists of items that, on average, are either of higher or lower difficulty than those in the first module. Student performance on all operational (i.e., non-pretest) items across both modules is used to calculate the section score.

Items from all four content domains appear in each test module. Across each module, items are arranged from easiest to hardest, allowing each test taker the best opportunity to demonstrate what they know and can do. Embedded pretest items appear in differing locations throughout the sequence.

Students may navigate freely through a given module of items; this allows them to preview upcoming items within the module or flag others to return to should time permit. However, once Bluebook has moved them to the second module in the Math section, students may not return to the items in the first module, nor may students return to the Reading and Writing section once Bluebook has moved them to the Math section.

### Test Length by Number of Items

The Math section consists of 44 items, which are divided into two equal-length modules, one for each of the section's two stages. Of each module's 22 items, 20 are *operational*, meaning that students' performance on them contributes to their section score, and 2 items are *pretest*. Answers to pretest items do not contribute in any way to students' Math section score; rather, their purpose is to collect student performance data that will be used by College Board to help assess whether these items are suitable for operational use at a future time. To students, pretest items appear the same as operational items, thereby ensuring students' maximal effort on the former. The number of pretest items embedded in the Math section is minimized to limit the impact of answering nonoperational items on students' test taking.

### Time Per Section and Module

Students have a total of 70 minutes to complete the Math section. This time is divided equally between the two modules, meaning that students have 35 minutes to answer the items in a given module. Once time has expired for the first module, students are automatically advanced to the appropriate (higher- or lower-difficulty) second module and may not return to the items in the first module.

### Average Time Per Item

Students have, on average, 1.59 minutes to answer each Math test item.

### Score Reported

Students receive a section score based on their overall performance on the Math section. This score is scaled from 200–800 for the SAT, 160–760 for the PSAT/NMSQT and PSAT 10, and 120–720 for the PSAT 8/9, each in 10-point intervals. This section score is added to the Reading and Writing section score to determine students' total score for the assessment.

### Item Formats Used

Approximately three-quarters of the Math items use the four-option multiple-choice format with a single best answer known as the *keyed response* or the *key*. The remaining items are in the student-produced response, or SPR, format. For this latter format, students must generate their own answers to the items and then enter their responses in Bluebook. Unlike Math multiple-choice items, for which only one correct response is provided among the answer choices, Math SPR items may have more than one answer that students could enter and have counted as correct, although they are directed to provide only one answer per item.

### Context Topics

Approximately 30% of Math items are set in context. These in-context ("word") items require students to consider a science, social studies, or real-world scenario and apply their math skills and knowledge, along with an understanding of the context, to determine the answer to each. To reduce the impact that topic selection might have on answering these items, contexts are developed that are either widely familiar or otherwise accessible to all students because of their grounding in common rigorous academic subject matter. Topic-specific prior knowledge is not assessed.

### Word Count

A majority of math-in-context items are 50 standardized (six-character) words or fewer in length. (For a discussion of how a standardized word count is obtained, see "Word Count" in Section 2.2.1, The Reading and Writing Section.)

### Informational Graphics

Select questions in the Math section are accompanied by an informational graphic. The inclusion of such graphics achieves two main goals: first, it authentically reflects the prominence and utility of such data displays in the field of math; second, it realistically assesses students' ability to locate, interpret, and use information from graphics to solve problems. Informational graphics in the Math section can take many forms, including, but not necessarily limited to, graphs of functions in the *xy*-plane, dot plots, scatterplots, bar graphs, line graphs, histograms, and representations of geometric figures.

### Text Complexity

Text complexity is not formally measured in the Math section. For the roughly 70% of Math items without a context, text complexity is irrelevant. For those items in context, the test development goals are to minimize the impact of linguistic factors on students' ability to answer the items while still presenting scenarios rich enough to support problem-solving in science, social studies, and real-world settings. To these ends, word counts are constrained, and the language used is kept as simple, clear, and direct as possible. Moreover, Math (and Reading and Writing) test development staff have been trained in the principles of linguistic modification (Abedi & Sato, 2008), a set of approaches that seeks to reduce language burdens in test items without altering the construct being measured or reducing intended rigor.

## Items

Math items come in two formats: four-option multiple-choice items, each with a single correct response among the answer choices, and student-produced response (SPR) items, wherein students must generate and enter their own answers to items and where there may be more than one correct answer (although students are directed to enter only one). The correct answer (potentially more than one for SPR items) is known as the *keyed response* or the *key*, while the three alternative answer options in multiple-choice items are known as *distractors*. Each multiple-choice distractor represents a common error that students might reasonably make in answering the item, so distractors are intended to be at least surface plausible; however, no distractor is meant to compete with the key as the best response for students with the targeted level of math achievement.

## Variations by Testing Program

The content of the Math section varies to some extent by testing program in order to reflect age and attainment differences across the test-taking populations. The following are of particular note:

- Rational and radical equations (Advanced Math) are not represented on PSAT 8/9.
- Trigonometry skills and knowledge are not assessed on PSAT 8/9.
- Skills and knowledge associated with circles (Geometry and Trigonometry) are assessed only on the SAT.
- In terms of number and proportion of questions, Algebra is most prominent in PSAT 8/9 and decreases slightly at higher program levels; the weighting of Advanced Math increases by program level; the weighting of Problem-Solving and Data Analysis decreases slightly by program level; and the weighting of Geometry and Trigonometry/Geometry remains largely consistent by level.

Other small variations by skill/knowledge testing point can be found in Appendix A.

**IMPORTANT**

This subsection describes the digital SAT Essay, an optional component of the digital SAT assessment. Presently, the digital SAT Essay is administered only as part of select U.S. school day testing in states that have contracted with College Board for this component. It is not currently offered as a student option on weekend digital SAT testing. Mode aside, the digital version of the SAT Essay is the same as its paper and pencil predecessor.

## 2.2.3 SAT Essay

The digital SAT Essay is designed to measure students' attainment of critical college and career readiness prerequisites in reading, analysis, and writing. The Essay is a single 50-minute direct-writing assessment during which students are presented with an argument that they must comprehend, analyze, and discuss in their written response. The Essay prompt, which is consistent from administration to administration apart from minor wording variations, requires students to explain how the author of the accompanying passage builds an argument to persuade an audience. This argument is written for a broad audience (e.g., an op-ed in a major newspaper) and requires no topic-specific background knowledge to read and understand. As they analyze and write about the passage, students are asked to consider how the author uses such elements as evidence, reasoning, and/or stylistic and persuasive elements to influence the audience's thinking.

The Essay is not designed to elicit test takers' subjective opinions. Instead of asking them simply to emulate the form of evidence used by drawing on their own experiences or imaginations, the Essay requires test takers to make purposeful, substantive use of textual evidence in a way that can be objectively evaluated. In keeping with the test's emphasis on relevant knowledge that students will continue to encounter throughout their education, this task is designed to promote the practice of reading a wide variety of arguments and analyzing how authors do their work as writers. Students' responses to the Essay prompt are evaluated along three dimensions—reading, analysis, and writing—as discussed in more detail below.

### Construct

The construct (concept) intended to be measured by the digital SAT Essay is reading, analysis, and writing achievement relative to core college and career readiness requirements in English language arts as well as in a range of academic disciplines.

### Claim

The SAT Essay is designed to elicit evidence from student test performance in support of the following broad claim about students' literacy achievement:

> *Students must be able to demonstrate college and career readiness proficiency in producing a cogent and clear written analysis using evidence drawn from an appropriately challenging source text written for a broad audience.*

## Key Features

Table 2.4 summarizes many of the key features of the digital SAT Essay. These features are discussed following the table.

**TABLE 2.4** DIGITAL SAT ESSAY: HIGH-LEVEL SPECIFICATIONS

| Characteristic | Digital SAT Essay |
|---|---|
| Administration | Single direct-writing task |
| Task length | 1 prompt |
| Time per task | 50 minutes |
| Scores reported | Three Essay dimension scores, each on a 2–8 scale:<br>Reading<br>Analysis<br>Writing<br>Scores are reported separately from each other and are not reflected in any way in the digital SAT's total or section scores. |
| Task format used | Single direct-writing prompt, consistent from administration to administration |
| Passage subject area | Arguments written for a broad audience |
| Word count | 750–800 (six-character) words per passage |
| Informational graphics | None |
| Text complexity band | High school (grades 9–12) |

The following discusses the key features identified in Table 2.4.

### Administration

The digital SAT Essay is administered as a single direct-writing task.

### Task Length

The digital SAT Essay consists of a single direct-writing prompt.

### Time Per Task

Students have 50 minutes to complete the digital SAT Essay task, which involves reading and analyzing an included passage and drafting a written response in accordance with the prompt.

### Scores Reported

Student responses to the digital SAT Essay task are scored and reported out on three *dimensions*: Reading, Analysis, and Writing. Scores on each dimension are scaled from 2–8. These individual dimension scores are not combined with each other, and they are not combined in any way with the section and total scores yielded by the other portions of the digital SAT.

Table 2.5 summarizes the criteria evaluated in generating digital SAT Essay scores.

**TABLE 2.5** DIGITAL SAT ESSAY: SCORING DIMENSIONS AND EVALUATION CRITERIA

| Digital SAT Essay Scoring Dimension | Evaluation Criteria |
|---|---|
| Reading | ▪ Comprehension of the passage<br>▪ Understanding of central ideas, important details, and their interrelationship<br>▪ Accuracy in representation of the passage (i.e., no errors of fact or interpretation introduced)<br>▪ Use of textual evidence (quotations, paraphrases, or both) to demonstrate understanding of the passage |
| Analysis | ▪ Analysis of the passage and demonstrated understanding of the analytical task<br>▪ Evaluation of the author's use of evidence, reasoning, and/or stylistic and persuasive elements, and/or features chosen by the student<br>▪ Support for claims or points made in the response<br>▪ Focus on features of the passage most relevant to addressing the task |
| Writing | ▪ Use of a central claim<br>▪ Use of effective organization and progression of ideas<br>▪ Use of varied sentence structures<br>▪ Employment of precise word choice<br>▪ Maintenance of a consistent, appropriate style and tone<br>▪ Command of the conventions of Standard English |

### Task Format Used

The digital SAT Essay is a single direct-writing task. While the passage students are asked to analyze changes, the task is always to explain how the passage's author builds an argument to persuade an audience.

### Passage Subject Area

Passages presented for analysis in the digital SAT Essay are arguments written for a broad audience for which permission to use has been obtained if necessary (i.e., not for works in the public domain). These passages are excerpted or (more typically) adapted in minor ways from high-quality previously published sources, generally editorials, op-eds, and the like. No topic-specific background information is needed to read and comprehend the passages; when necessary, small bits of contextual information, such as an advance organizer or footnotes, are supplied.

### Word Count

Each digital SAT Essay prompt includes a passage that students must read and analyze. This passage ranges from 750 to 800 standardized (six-character) words. (For a discussion of how a standardized word count is obtained, see "Word Count" in Section 2.2.1, The Reading and Writing Section.)

### Informational Graphics

Passages used in the digital SAT Essay do not include informational graphics.

### Text Complexity Band

Digital SAT Essay passages are drawn from the high school (grades 9–12) text complexity range. As in the Reading and Writing section, text complexity for Essay passages is determined using a combination of quantitative and qualitative measures.

## Task

Task format is unvarying across all administrations of the digital SAT Essay, with the exception that the passage students need to analyze varies (along with minor prompt wording variations accounting for the differing passages). Each Essay task includes the following elements:

- Introductory directions indicating that students should consider how the author of the passage uses evidence, reasoning, and/or stylistic and persuasive elements
- The passage students are to analyze
- The prompt proper, which, in generic terms, appears as follows, with variable text appearing in [brackets]:

Write an essay in which you explain how [the author] builds an argument to persuade [their] audience that [author's claim]. In your essay, analyze how [the author] uses one or more of the features listed above (or features of your own choice) to strengthen the logic and persuasiveness of [their] argument. Be sure that your analysis focuses on the most relevant aspects of the passage.

Your essay should not explain whether you agree with [the author's] claims, but rather explain how the author builds an argument to persuade [their] audience.

## 2.3 Statistical Specifications

The assembly of a test is governed by both content and statistical specifications, which together compose the test specifications. This subsection describes the statistical specifications applied in the test assembly process for the digital SAT Suite.

The digital SAT Suite assessments are developed according to a multistage adaptive testing (MST) model. Student testing in the Reading and Writing and Math sections is conducted in two stages. In the initial (routing) stage, students respond to a collection of items (known as a *module*) across a broad spectrum of difficulty; their performance on the first module of items determines whether they receive a second-stage module composed of items that are either, on average, of higher or lower difficulty than those in the initial module, though both second-stage modules contain easy, medium-difficulty, and hard questions, albeit in differing proportions.

Due to their adaptive nature, the digital SAT Suite tests are significantly shorter than the paper-based tests they have replaced. Therefore, a key consideration in establishing statistical specifications is to maintain or exceed the score reliability of the paper-based SAT Suite tests. In addition, it is important to ensure accuracy in routing students to the second stage of testing in each section. Within and across administration windows, it is critical to ensure the uniformity of psychometric properties for all test instances taken by students as well as to minimize potential gaps in terms of reported scores (i.e., it should be possible for all valid score points to be attained). Last but not least, statistical specifications need to support the sustainable capacity of the item pool to build comparable tests in the foreseeable future.

The digital SAT Suite tests are based on item response theory (IRT), in the sense that the tests are constructed and scored with item parameters estimated from pretesting. The statistical specifications are, as a result, articulated in terms of IRT test information functions (TIFs) and test characteristic curves (TCCs).

TIFs are used within an IRT framework and are related to the concepts of reliability and standard error of measurement. A key requirement of score comparability is that the various test panels or linear forms of the assessment have similar estimates of score reliability and, in the case of digital SAT Suite, maintain or exceed the reliabilities of the corresponding paper-based SAT Suite tests.

In the context of IRT, College Board achieves these goals by requiring that the TIFs meet a specific target at various points on the ability (IRT theta) continuum. TIF targets are established for each testing program (digital SAT, PSAT/NMSQT and PSAT 10, and PSAT 8/9) and for each type of module (initial [routing], higher difficulty, lower difficulty). The values of TIF targets are specified such that the overall score reliability of the assembled tests can be maintained at the desired level no matter which route (initial to higher difficulty, initial to lower difficulty) a given student may take.

Two constraints are placed on these TIFs. First, the automated test assembly (ATA) software College Board uses must minimize the difference between the target TIF and the TIF estimated from the items being selected. In reality, the differences between the TIFs will never be zero, so the second constraint places a bound around the estimated TIF, which requires that the difference between the target TIF and the estimated TIF be as small as possible but no larger than the specified range. This range is based on internal trial builds showing that scores arising from test panels meeting these constraints result in scores that have similar reliability.

To assist in the specification of TIF targets, a reliability estimate for the digital SAT Suite tests is needed. Within an IRT-based framework and scoring procedures that treat the parameters of each item as fixed, reliability could be calculated in a few different ways. Given the fact that the digital SAT Suite tests are adaptive, College Board has found using the ratio of the standard error of the theta (Andrich, 1988) to observed theta to be most appropriate. This is because by simulating examinees through an MST, College Board can account for the adaptive structure of the test when computing reliability using the following formula:

$$Reliability = 1 - \frac{1}{\theta_X^2} \sum \left[ \theta_E^2 x\ Slope^2\ x\ Weight \right]$$

The theta standard errors can be transformed into scale score standard errors by multiplying them by the slope of the theta-to-scale-score transformation. Then the average of the squared scale score standard errors is divided by the variance of the scale scores and subtracted from 1.

Another consideration for TIF target specification is where to place the cut score for the routing module. There are various strategies, such as placing it near or around an important score point, such as the College and Career Readiness Benchmark. An alternative method is having around 50% of the exam population routed to either the lower- or higher-difficulty modules. To do this, College Board can use historical data to predict expectations for future ability distributions and set the cut score accordingly. In other words, College Board used the median test scores from archival data to set the cut score for the routing module.

The constraint on TCCs is analogous to the traditional concept of equating. The digital SAT Suite is pre-equated, meaning that equating after an administration is not performed and that all equating is done at the time of pretesting calibrations and test assembly.

The expected TCCs for the digital SAT Suite tests are set by testing program and type of module, and they are minimized toward a target but have a fixed range within which all TCCs for the same module type must fall. These ranges were based on research establishing that all test panels meeting this requirement will produce similar score distributions, meaning that it would be a matter of indifference to students as to which panel they were assigned. As stated earlier, a side benefit of specifying appropriate TCC targets across the ability distribution is minimizing potential gaps in the reported scale scores. Such gaps result when certain ranges of ability estimates are either impossible or hard to obtain by the expected student populations.

Each test panel is built to the same content specifications as all others, ensuring that content coverage is consistent. To maintain statistical uniformity, College Board chose to constrain the TCC. By the principle of a "difference that matters" (Dorans & Feigenbaum, 1994), a range of ±0.5 of the expected true score was used as a threshold for the TCC target. That is to say, tests are built to within ±0.5 of the expected true score of the TCC target. Because targets for the digital SAT Suite's MST model are set at the level of the module, of which there are two per selected-response section, each module is set up with a threshold of ±0.25 of the expected true score. Therefore, when the two modules of a section are administered together in a single exam, the maximum possible deviation from the TCC target is ±0.5 of the expected true score. The combination of a hard target for the TCC and a soft target for the TIF allows for greater degrees of freedom for the ATA software while allowing adequate control of the uniformity of the exam.

In addition to the MST forms described above, College Board develops and administers linear (nonadaptive) versions of the digital SAT Suite tests. Such forms are intended only for students requiring paper-based testing or supplemental materials for accommodations or circumstantial reasons, and, except for being longer than their digital counterparts due to their nonadaptive nature, these linear tests conform to the same specifications as those used in creating MST forms. The above principles regarding the TIF and TCC targets used to assemble MST panels were also applied to the assembly of linear forms, with the exceptions that no distinction was made between module types and that the targets were adjusted to correspond to the extended linear test length. As a result, the reliability of linear digital SAT Suite scores is comparable to that of its MST tests, and the statistical qualities of the linear tests are consistent over time.

Chapter 3

# Test Development and Assembly

## 3.0 Introduction

This chapter details the development and assembly process used in the creation and delivery of the tests of the digital SAT Suite. The test development and assembly processes for the digital Reading and Writing and Math sections as well as for the digital SAT Essay are first discussed, followed by the approaches used to manage the resultant inventory of test materials and to assemble test panels. Section 3.1, Test Content Development, acquaints readers with the rigorous approach College Board uses to develop test items for the digital suite. Section 3.2, Inventory Management, discusses how the created items are stored and maintained in College Board's item bank system. Section 3.3, Test Assembly, describes how items are assembled into multistage adaptive and linear tests.

## 3.1 Test Content Development

This section provides an overview of the processes used in producing test items for the digital SAT Suite assessments. The main purpose of this section is to acquaint readers with the rigorous approach College Board uses to develop items for the digital suite and thereby provide greater transparency into a critical aspect of delivering the digital SAT Suite assessments in a valid, reliable, and fair way. The Reading and Writing and Math section discussions are combined, as they are conceptually similar, with a discussion of the digital SAT Essay following.

### 3.1.1 The Reading and Writing and Math Sections

#### Digital SAT Suite Automated Item Generation (AIG)

Before examining the procedural steps involved in creating test items for the digital SAT Suite Reading and Writing and Math sections, it is critical to understand the basics of the test development paradigm used for the suite.

As part of the transition from paper-based to digital SAT Suites, College Board made significant investments in technology supporting not only test delivery but also the development and deployment of assessment content. In particular, College Board invested in technologies to facilitate automated item generation, or AIG (Drasgow et al., 2006; Gierl et al., 2008; Fu et al., 2022). AIG as employed by College Board is a "best of both worlds" approach: the process relies heavily on the skill of trained human experts while leveraging technology to greatly increase item yields and thereby enhance test security. This stands in contrast to traditional item-writing methods, which typically involve test developers producing bespoke test items one at a time.

The fundamental unit of Reading and Writing and Math test development in the digital SAT Suite paradigm is the *parent model*. A parent model is, in some sense, the "base" form of a given test item, though, strictly speaking, the parent model is not itself an item but rather a template carefully crafted by content experts from which a series of individual test items, known as *child items*, are generated. Child items for any given parent model differ from each other through the inclusion of *variable elements*, or *variables*.

These variables are intended to be *incidental*, meaning that their alteration does not affect the underlying substance, structure, or logic of the "base" item or the performance characteristics of the resultant child items. When only incidental variables with appropriate values are used, any student receiving any one of the child items from a given parent model should have the same experience as any other student receiving a different child item from the same parent model, a hypothesis that is confirmed or refuted for any given parent model via pretesting.

The permissible range of values for each of the variable elements in any one parent model is defined at the time of the model's development and potentially refined through the authoring and review process. This collection of values is known as a *dataset* in Reading and Writing test development, and is defined by *parameters* in Math test development.

The child items generated from any given parent model are intended to be *isomorphic* in the sense that each child item generated from that parent model is expected to perform statistically similarly within a very narrow performance window: ±0.05 of the average IRT *p*-value of the child items. Child items that are verified to be isomorphic are considered functionally identical and can be used interchangeably, thereby enhancing the variety and security of the resultant item pool.

The *isomorphic hypothesis*—that is, the claim that the child items from a given parent model are sufficiently similar in performance to be used interchangeably—is verified or rejected through pretesting. A representative subset of the child items from each parent model is pretested on a large sample of the digital SAT Suite test-taking population and the performance of the child items assessed. Should all pretested child items perform similarly—within ±0.05 of the pretested child items' average IRT *p*-value—the isomorphic hypothesis is deemed proved, and, barring any other issues with the pretested items (such as evidence of differential item functioning, or DIF, discussed in Chapter 6, Section 6.4, Item Analysis, Calibrations, and Pre-Equating), all child items produced from the model are considered eligible for use in operational testing. Should sufficient variance arise in how the child items performed in pretesting, the isomorphic hypothesis is considered disproved; the entire parent model and all its child items are excluded from the operational item pool, though the model may, at the discretion of College Board subject matter and measurement experts, be re-edited and a subset of its (revised) child items re-pretested.

The values that populate Reading and Writing datasets and that fall within ranges defined by Math parameters, like all aspects of digital SAT Suite test items, are carefully researched and vetted. For all Reading and Writing parent models and for Math parent models set in context, these values are evaluated to ensure that the information they represent is accurate (in cases where this is relevant) or plausible (in cases where this is not, such as literature) as well as appropriate for the audience and the occasion of testing. All Math parameters, whether for items in or out of context, are further evaluated to ensure that the permissible values are limited to those that are likely to engender the same basic level of computational effort across child items.

The parent model/child item concept employed in digital SAT Suite Reading and Writing and Math test development is depicted schematically in Figure 3.1.

**FIGURE 3.1** DIGITAL SAT SUITE AIG DEVELOPMENT PARADIGM

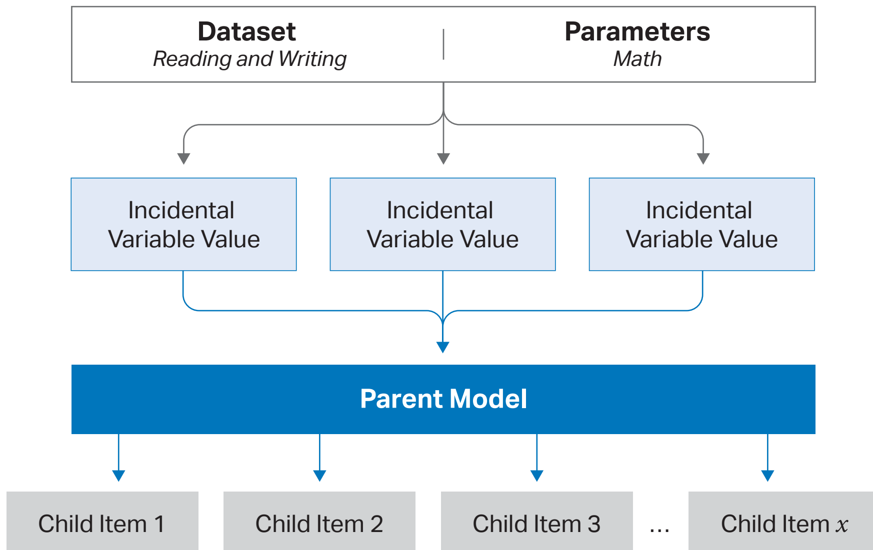



Figure 3.1 depicts how a dataset in Reading and Writing or the parameters in Math feed into the associated parent model by altering the values of incidental variables across child items. This process yields a series of child items, each of which is highly similar in substance but differing in superficial ways to produce item variety. In theory, a parent model could produce a number of questions limited only by the size and complexity of the Reading and Writing dataset or the breadth of the range of the Math parameters, hence the designation "Child Item *x*" in the figure; in practice, however, the number of child items produced from an individual parent model is intentionally kept relatively low so as to preclude the resultant item pools from becoming overly homogenized and predictable.

## Item Development Process

### Item Conception

College Board subject matter experts begin the test development process by conceptualizing each item to be written in accordance with the tests' specifications and their own development assignment. These assignments identify high-level characteristics of the items to be produced, such as the skill/knowledge element to be assessed (e.g., in Reading and Writing, the use of textual evidence); the subject area, if any, in which the skill/knowledge element should be assessed; the targeted text complexity of the passage, if applicable (e.g., a reading level equivalent to that commonly expected of students in grades 9–11); and the intended difficulty of the resultant child items. Each assignment is conceptualized at the parent model level, with resultant child items each required to conform to these specifications.

## Topic Selection, Vetting, and Research

For all Reading and Writing development and for the development of in-context items in Math, College Board subject matter experts begin parent model development by researching potential stimulus topics that align with the assignment; seem likely to yield numerous child items that are psychometrically, cognitively, and experientially equivalent; and present different students with highly comparable testing experiences. Potential topics (e.g., economists' explanations for various import/export patterns around the world, a study about the nutritional needs of bears) are vetted by College Board test development staff to ensure high-level appropriateness for the student audience and the occasion of testing. As part of this process, these experts may consult the detailed digital-suite test specifications, empirical research underlying the specifications, fairness guidelines, additional research on the topic in question, commonly used textbooks at the targeted grade level in the relevant subject area, studies identifying essential prerequisite skills for relevant entry-level college courses, performance data from previously developed items, and other resources.

## Parent Model and Dataset/Parameter Development

Once a topic and task have been identified and vetted, content experts craft the parent model. When developing parent models, these experts consult detailed digital-suite test specifications to ensure that the task presented in the model is aligned with the evidence base corresponding to the item type. In addition to documented criteria for item soundness, validity, and fairness, content experts consult internal documentation on the production of isomorphic (that is, superficially distinct but psychometrically, cognitively, and experientially equivalent) child items. This documentation, which is heavily informed by a robust body of academic literature on template-based item development as well as by years of test development experience, helps guide content experts in selecting the portions of a parent model to vary across child items.

In Math items, these variable elements are defined by carefully determined parameters set by content experts. These experts work to ensure these values are set as to result in comparable cognitive loads for students receiving any of the child items. In Reading and Writing items, all acceptable variable values are preidentified by content experts. In either case, variable elements are selected and crafted to have no effect on students' performance on the resulting child items and to be fair to all test takers.

Additionally, when relevant, variable elements are evaluated to ensure that resultant child items make accurate or, in cases where items are posing hypothetical scenarios, plausible real-world statements. For example, the parameters for a parent model about the number of calories per day that brown bears need to consume would not yield a child item that says that these bears need only 350 calories daily—an unreasonably low number—but would instead yield child items that present values in the real-world range of 5,000 to 20,000 calories per day.

### Child Item Generation

After the initial parent model conceptualization and development by College Board test development staff is complete, the parent model, including its variable elements and associated values, is loaded into College Board's item development platform to generate child items. As previously noted, this process populates the resultant child items with the incidental values defined in the dataset/parameters while keeping the overall "frame" of the parent model intact. Throughout the process of parent model development, content experts may iterate on the model to improve its overall quality or to better ensure the isomorphism of the resultant child items.

### Internal and External Review

Once a parent model has been created, a robust internal review process ensures that the model and all yielded child items meet rigorous quality standards.

Multiple College Board content experts review each parent model, all variable content (variable elements and dataset/parameter values), and the generated child items. Such reviews ensure that the items are aligned to the intended constructs, key affirmatively, are free from fairness problems, and are appropriately targeted to the intended difficulty level and to the educational attainment of the test-taking population. Parent model developers revise content in response to review feedback until the model meets documented standards.

Additionally, the generated child items receive multiple rounds of review by College Board's editorial and accessibility teams to ensure that the items are free from errors and render correctly in the digital test and in alternate formats.

College Board also routinely engages independent experts—primarily active teachers at the secondary and postsecondary levels—in two related ways to assess the content soundness and fairness of digital SAT Suite test materials. First, College Board retains diverse teams of external experts to evaluate the fairness of all Reading and Writing items as well as all Math items set in context, as these are the items most likely to inadvertently introduce fairness concerns. Second, College Board hosts semiannual external reviews of representative test materials with groups of independent subject matter and fairness experts. These review committees consist of diverse groups of educators at high schools, community colleges, and public and private four-year colleges from across the United States; collectively, reviewers' credentials include both subject matter and fairness expertise, with the latter addressing matters of fairness in general as well as fairness as it pertains to Black/African American, Latino, Asian American, and American Indian/Native American students, students from all genders, students who are English learners, and students with disabilities. For these reviews, sample digital SAT Suite test forms/panels are built from the operational item pool by the same test assembly program that generates operational test forms/panels for students, thus ensuring that the reviewed materials adhere to the suite's specifications and are authentic and fair representations of what students could encounter on test day.

The primary purposes of these semiannual reviews are, first, to solicit feedback from independent content and fairness experts on the alignment of test materials to the suite's testing purposes and specifications as well as to postsecondary readiness prerequisites and secondary standards and curricular emphases and, second, to solicit item-level feedback with respect to content and fairness. Jointly, the reviews both identify problematic test content and help to guide future test development in productive ways.

## Pretesting

For the digital SAT Suite Reading and Writing and Math sections, College Board employs an embedded pretesting model. In *embedded pretesting*, a small number of nonoperational (unscored) items are included among the operational ones presented to students to obtain the performance statistics necessary to evaluate the suitability of the former for use on future tests. These pretest items are *embedded* in the sense that they appear among and are indistinguishable to the test taker from the operational items they receive. Embedding helps ensure that test takers give pretest items their full attention and effort, while strictly limiting the number of pretest items given to each test taker minimizes the impact of answering items that do not contribute to their scores.

The Reading and Writing and Math sections each contain four slots for pretesting: two in the initial (routing) module and two in each second-stage (higher- and lower-difficulty) module. Any given student taking one of the digital SAT Suite tests will see a total of eight pretest items, four in each test section. The pretest items appearing in the second-stage modules of each panel are the same so that the items are pretested across the full range of student achievement. Pretest items are randomly assigned, subject to some content considerations, to a placement within each module to reduce location effects on the resulting item statistics and to diminish the likelihood of students being able to distinguish pretest from operational items.

For each parent model, a representative subset of its child items is pretested rather than the full set, thereby increasing pretesting efficiency and allowing for the generation of item pools larger than would be possible if all child items from each parent model were pretested. This is possible because, as previously noted, each child is assumed to be isomorphic during development and verified as such via pretesting. Should the isomorphic hypothesis hold true via pretesting, each child item from a given parent model, whether or not it was itself pretested, takes on the statistical characteristics of the parent model, which are derived from the statistics of the pretested child item subset. Differential item functioning (DIF) analysis is also performed at the parent model level to examine whether, in statistical terms, the child items favor or disfavor one or more defined population subgroups (e.g., students identifying as female relative to students identifying as male) based on an analysis of samples matched in terms of achievement on the construct of interest. (For more details on pretest and DIF analysis for the digital SAT Suite, as well as a fuller discussion of the pretesting process, see Chapter 6, Psychometrics.

## 3.1.2 The Digital SAT Essay

In the digital SAT Essay, College Board has developed a rich task that authentically reflects the work students need to do to be ready for and successful in college and careers. At present, the Essay is offered only as a part of select U.S. state assessment contracts in which all students are given the task as part of school day testing.

The aim of the digital SAT Essay is to determine whether students can demonstrate college and career readiness proficiency in reading, analysis, and writing by demonstrating comprehension of a high-quality source passage and producing a clear and cogent written rhetorical analysis of how that passage builds an argument to persuade an audience. Although the Essay source passage varies from administration to administration, the Essay task itself and the prompt language is highly consistent. Such consistency allows students, in their preparation and during the actual test, to focus squarely on the source analysis task.

## Crafting of Passages and Prompts

All Essay passages are excerpted or minimally adapted from high-quality, previously published sources for which permission for use has been obtained from their respective copyright holders, if any. While the specific style and content of the passages vary across the Essay item pool, the passages take the general form of what might be called *arguments written for a broad audience*. The passages examine ideas, debates, trends, and the like in the arts, the sciences, and civic, cultural, and political life that have wide interest, relevance, and accessibility to a general readership. Essay passages are also selected based on their use of evidence, logical reasoning, and/or stylistic or persuasive elements so that students have the best opportunity to demonstrate their ability to analyze how the author built the argument. The text complexity of the passages is carefully monitored to ensure that the reading challenge is appropriate and comparable across administrations but not an insurmountable barrier to test takers responding to the passage under timed conditions. Prior knowledge of the topics of the passages is neither expected nor required.

As much as possible, Essay passages are kept as they originally appeared in publication. In other words, College Board avoids, or limits as much as possible, the editing or excising of portions of a passage. This allows the passage to maintain the integrity of the argument as the author originally wrote it. Only limited edits or deletions are made for the purpose of word length, fairness, or obscurity/difficulty. When elisions or minor additions are made, they are indicated by ellipses or brackets in order to maintain transparency for readers. Any additional editing or adaptation beyond minor changes risks further compromising the passage as it originally appeared.

College Board prompt developers are trained to locate passages suitable for use on the digital SAT Essay. As part of the passage-finding process, College Board staff selects passages, which are then either accepted or rejected for further development based on discussions and evaluations by the Essay development team. Once Essay passages have been accepted, College Board crafts a prompt-specific summary statement for each passage. This statement comes at the end of each passage in the prompt text that outlines the Essay task. The prompt summary statement provides test takers with the main claim of the passage so that they can focus on demonstrating a more detailed understanding of the passage and on analyzing how the author built the argument.

## Passage Content and Fairness Reviews

Once passages have been selected that meet internal criteria for topic, length, text complexity, and appropriateness for the task and occasion, independent external reviewers, chiefly secondary and postsecondary classroom teachers, review the prompts for any issues of fairness, as defined by documented College Board criteria. Prompts are reviewed to ensure that they are broadly accessible to a diverse audience of test takers, allow test takers to demonstrate their achievement on the construct, and do not advantage or disadvantage any particular subgroup in the student population based on construct-irrelevant factors such as race/ethnicity, gender and sexual identities, or socioeconomic status.

Because an argument by its very nature must be debatable to some degree, the subject matter of digital SAT Essay passages occasionally falls into potentially sensitive territory. Topics for the Essay may sometimes be perceived as sensitive because they encompass rich and complex issues in which differing perspectives may sometimes come into conflict. However, College Board is careful to select topics that are appropriately subject to debate and yet unlikely to cause an emotional response for students. Furthermore, students are told explicitly in the Essay directions that the task is not about their personal views on the topics represented, and the inclusion of such opinions is neither rewarded nor penalized in the scoring process (although test takers who devote significant attention to offering their own views are likely hampering their performance by being off-task).

In addition to the initial fairness review of digital SAT Essay passages, College Board implements a second round of internal checks on passage suitability prior to making assignments of Essay prompts to particular digital SAT test administrations. This review is intended to confirm that each proposed passage, which was developed and banked in prior years, remains acceptable for use. A previously vetted passage may still, at this stage, be determined not to be usable in operational testing due to changes in the political and social climate making formerly uncontroversial passages newly problematic or sensitive.

## Prompt Field Testing and Analysis

All digital SAT Essay prompts are put through rigorous field testing to ensure their suitability and comparability. As a group, the students who participate in field testing vary by race/ethnicity, gender, and socioeconomic status. They come from rural, suburban, and urban populations and attend both private and public schools. Once a sufficient pretesting volume has been reached, College Board oversees the scoring of Essay responses according to the digital SAT Essay rubric. The procedures used to score field test responses are the same used to score the Essay after it is administered operationally to test takers. Each response is scored independently by two raters, who assign a score of 1–4 in each of three dimensions: Reading, Analysis, and Writing. If the two independent rater scores on each dimension are in perfect agreement or are adjacent to each other (i.e., differ by exactly one point), the scores from the two independent raters are added together to give test takers a 2–8 score on each dimension. If the two independent raters' scores on one or more dimensions are nonadjacent (e.g., scores of 1 and 4), the response is rated by a third, senior-level rater. The third rater's score is then doubled to give test takers a 2–8 score on the relevant dimension(s).

To ensure fair and comparable digital SAT Essay prompts, College Board completes psychometric analyses on the field test results of all prompts, including analyzing the mean and frequency distribution by score point across each dimension for each prompt and evaluating interrater reliability stats (including correlations) between rater 1 and rater 2 (and rater 3 if applicable) for all dimensions (Reading, Analysis, Writing). The performance of field-tested prompts is analyzed across all prompts field-tested within a given year; their performance is also compared year over year to ensure mean and demographic comparability for each testing cycle. Any prompt that is an outlier—i.e., that has a mean score on any dimension that is unusually high or low or that has a distribution of scores on any dimension that is significantly different from the distribution of scores on the other dimensions—is not used operationally. Demographic analyses are also conducted, including analyzing female/male mean differences by dimension scores by prompt and mean differences by race/ethnicity for dimension scores by prompt. Any prompt that demonstrates extreme differential performance, either between males and females or between Whites and members of other races/ethnicities, is not used operationally.

## 3.2 Inventory Management

Once digital SAT Suite items, including SAT Essay prompts, are developed and pretested on a representative sample of the population of interest, they are officially stored and maintained in College Board's item bank system. The item bank system contains all items developed for each test of the digital SAT Suite coded by their item classifications, as specified by the test specifications. For all items and stimuli, the bank holds text, art/graphics, item codes for required elements, and statistical records for appropriate test administrations. Additional information held in the item bank (e.g., item status and usage) allows the appropriate items to be selected and placed into the test assembly item pool.

Items are sent from College Board authoring applications to the item bank once they are deemed ready for operational use. Prior versions of items are stored separately and are accessible if needed. Coding for item content (e.g., content domain, skill/knowledge element) and current item status (e.g., ready for pretesting) are controlled in the system and consumed by the automated test assembly software. Content usage rules (e.g., item is suspended) and tracking (e.g., item was used in a practice test) are thoroughly documented.

Items are developed for pretesting and later operational use based on an ongoing analysis of the operational pools, with development assignments targeting specific content and item types as appropriate. This ongoing pool analysis and replenishment informs item development plans and ensures the maintenance of robust operational pools that support content assembly needs for all test administrations. These operational item pools in Reading and Writing and in Math are sufficiently large that each student testing digitally is administered a highly comparable but unique test form, and all test forms administered meet all content and statistical specifications.

Test content is designed to support access to all students. The item bank is not currently certified as meeting all Accessible Portable Item Protocol® (APIP®) standards. However, it is able to support the majority of APIP requirements for accessibility. Exams are also packaged for delivery to students in such a manner that allows for extended time and extra breaks for students with timing accommodations.

## 3.3 Test Assembly

### 3.3.1 Multistage Adaptive Versions

Digital SAT Suite delivery operates primarily according to a multistage adaptive testing (MST) model. For the digital SAT Suite tests, MST involves administering to test takers Reading and Writing and Math test sections divided into two equal-length, separately timed stages, each consisting of a module of items. The primary output of test assembly is a series of test panels, each consisting of an initial (routing) module and associated lower- and higher-difficulty second-stage modules.

An individual student testing experience, known as a test form, consists of a panel's routing module and either (depending on student performance on the first module of items) the lower- or higher-difficulty second-stage module.

In College Board's implementation of MST, no initial ability estimates based on prior assessments are made. Item selection is performed by automated test assembly (ATA) software, which takes into account requirements related to the test's content specifications, statistical constraints, and item exposure limitations. For any MST panel to be used operationally, it must meet all specified constraints, which are designed to ensure score comparability as well as comparability across testing experiences. All MST panels are fixed length, and no additional termination criteria are used. Given the embedded design, pretesting is continuous throughout each digital administration. Items are developed for pretesting based on ongoing analysis of the operational pools, with assignments targeting specific content and item types as appropriate. This ongoing pool analysis and replenishment ensures the maintenance of robust operational pools that support the content assembly needs for all test administrations.

Operational adaptive test forms are assembled from a secure item pool composed of items that have not been released, disclosed, or otherwise made publicly available. Operational item pools in Reading and Writing and in Math are sufficiently large that each student testing digitally is administered a highly comparable but unique test form, greatly enhancing test security. No two students will be given the exact same items, removing the risk of a student gaining an unfair advantage by accessing test forms in advance or copying off their neighbors. In addition, students testing multiple times will receive content that is unique relative to any of their previous administrations.

### 3.3.2 Linear Versions

Linear (static) test forms for the digital SAT Suite are made available for the very small population of students who require paper-based testing (either a paper test booklet or paper components that accompany a static digital form) due to an accommodation need or religious exemption or for students testing in specialized facilities with regulations and/or security protocols that prevent online access.

The linear form is designed to parallel the statistical properties of the multistage test by having similar test information functions (TIF) specifications. The TIF of the linear form is designed to have as much information as the multistage test in the relevant areas. In the lower ability ranges, the TIF of the linear test parallels that of the lower-difficulty multistage portion; in the higher ability ranges, the TIF of the linear test parallels that of the higher-difficulty multistage portion. A given linear form is longer in terms of both number of items and time allotted to attain the requisite coverage but measures the same skills and knowledge in the same proportions as the corresponding multistage adaptive test.

Although the volume of students testing outside of the digital testing application is extremely low, the linear form inventory comprises a number of distinct test forms, enhancing test security as well as supporting unique test content for students who test multiple times.

Chapter 4

# Test Administration and Security

## 4.0 Introduction

This chapter documents how College Board administers the digital SAT Suite in a manner that ensures all test scores are valid for their intended use and that all test takers have a fair testing experience. It also covers steps taken to prohibit the inappropriate sharing of test information.

Section 4.1, Test Requirements, discusses the parameters and procedures used in the standardized administration process that enables students to experience digital SAT Suite tests under a uniform set of conditions. Section 4.2, Digital Testing Application, examines the test administration app itself, including the general benefits of digital testing, the grounding of the digital testing application in principles of universal design, the provision of accessibility features, such as universal tools for all students and accommodations and supports for students who require them to fairly access test content, and the availability of authentic test practice. Section 4.3, Form Assignment and Administration, takes a closer look at components of the digital SAT Suite's adaptive testing model, including the use of multistage testing, embedded pretesting, and discrete (standalone) test items. Section 4.4, Security Protocols and Analyses, addresses the critical issue of test security both conceptually and in terms of how it is maintained on the digital-suite tests.

## 4.1 Test Requirements

In keeping with best practices and professional standards in large-scale standardized assessment (AERA et al., 2014), College Board has established procedures and parameters to ensure that the digital SAT Suite is administered to all test takers in a fair, equitable, and standardized manner. The goal of standardized administration is to enable all test takers to experience a uniform set of conditions so that test scores from different administrations can be used interchangeably for reporting, counseling students, and making admission and placement decisions.

### 4.1.1 Administration

Each assessment of the digital SAT Suite is composed of two sections: a Reading and Writing section and a Math section.[5] For individual students, the Reading and Writing and the Math sections are, in turn, composed of two equal-length *stages* consisting of *modules* of test items: an initial (routing) module consisting of a broad mix of easy, medium-difficulty, and hard items and, depending on student performance on the first module, a second-stage module containing items that are of either higher or lower average difficulty. Because the tests are designed according to a two-stage adaptive model (see Chapter 3, Section 3.3.1, Multistage Adaptive Versions), each module is separately timed, and while students may navigate freely within each module, they may not return to the first module's items once they have transitioned to the second module nor return to the first section (Reading and Writing) after moving to the second (Math). Up-to-date administration procedures can be found on the digital SAT Suite website, **satsuite.collegeboard.org/digital**.

### 4.1.2 Test Length

Each Reading and Writing module consists of 27 items. Of the items in each module, 25 are operational, the answers to which contribute to students' scores; an additional 2 items per module are pretest items, performance on which does not affect students' scores and whose inclusion is designed to help determine the suitability of using these items in future tests. In total, the Reading and Writing section is made up of 54 items, 50 of which are operational and 4 of which are unscored pretest items. Each Math module consists of 20 operational items and 2 pretest items, for a total of 44 items (40 operational, 4 pretest) across each test form. Note that the same second-stage pretest items are presented to students taking either the higher- or lower-difficulty module, ensuring that these items are evaluated relative to the full range of student achievement. (For more details on embedded pretesting in the digital SAT Suite, see Section 4.3.2, Embedded Pretesting).

### 4.1.3 Time Per Module

Each Reading and Writing module is 32 minutes in length, while each Math module is 35 minutes. As noted above, each module is separately timed. When time runs out on the first module of each section, Bluebook automatically moves students to the second-stage module, where they are administered either the lower- or higher-difficulty module associated with the routing module. When students complete the Reading and Writing section, they are automatically moved to the Math section after a short break between the sections.

### 4.1.4 Total Number of Items

Each Reading and Writing section consists of 54 items (4 of which are pretest), while each Math section consists of 44 items (4 of which are, again, pretest items).

### 4.1.5 Total Time Allotted

Students have 64 minutes to complete the Reading and Writing section and 70 minutes to complete the Math section.

---

5 Select U.S. digital SAT school day administrations also include the Essay, a direct writing assessment. The digital SAT Essay is discussed in detail in Chapter 2, Test Specifications, and Chapter 3, Test Development and Assembly.

## 4.1.6 Average Time Per Item

Students have, on average, 1.19 minutes to answer each Reading and Writing item and 1.59 minutes to answer each Math item.

## 4.1.7 Scores Reported

Each digital SAT Suite assessment yields three scores:[6] a total score and two section scores. The total score is based on students' performance on the entire assessment and is the arithmetic sum of the two section scores. Two section scores, one for Reading and Writing and the other for Math, are based on students' performance on each section. (For more details on digital SAT Suite scores, see Chapter 5, Test Scoring and Reporting.)

## 4.1.8 Item Format(s) Used

The Reading and Writing section exclusively uses four-option multiple-choice items, with each item having a single best answer (known as the *keyed response* or *key*). Approximately 75% of items in the Math section use the same four-option multiple-choice format, while the remainder use the student-produced response (SPR) format. As the name implies, students answering the latter type of Math item must generate their own response and enter it into a field in Bluebook. These items assess students' ability to solve math problems with greater independence and with less structure and support than that provided in the multiple-choice format. SPR items may have more than one correct response, although students are directed to supply only one answer.

## 4.1.9 Stimulus Subject Areas

The digital SAT Suite assessments ground all Reading and Writing and some Math items in authentic contexts based in academic disciplines or real-world scenarios. In Reading and Writing, each of these contexts includes a brief passage (or pair of passages) as well as a single (discrete) question based on the passage(s). Reading and Writing passages are drawn from and reflect the norms and conventions of the subject areas of literature, history/social studies, the humanities, and science. Students do not need topic-specific prior knowledge to answer Reading and Writing items; all the information needed to answer each item is provided in the passage or passages themselves.

In Math, about 30% of test items are set in academic (science or social studies) or real-world contexts, while the rest are "pure" math items outside of context. Math contexts are brief: sufficient in length to establish a scenario but clear, direct, and concise enough not to impose undue linguistic burdens on students. Contexts set in science or social studies emulate the kinds of problems, reasoning, and solving strategies commonly encountered in those fields, adding to the tests' verisimilitude. Again, topic-specific prior knowledge is not required to answer these sorts of items.

6 Select U.S. SAT school day administrations that include the Essay direct writing assessment also yield Reading, Analysis, and Writing dimension scores, each on a 2–8 scale, based on student performance on the Essay task.

### 4.1.10 Informational Graphics

In accordance with evidence about essential college and career readiness requirements, both the Reading and Writing and Math sections include informational graphics with select items. For Reading and Writing, these informational graphics are restricted to tables, bar graphs, and line graphs, as these are the most commonly used types in academic and real-world texts; for Math, the range is wider and includes several types of data displays (e.g., scatterplots) as well as geometric figures and *xy*-plane graphs.

## 4.2 Digital Testing Application

### 4.2.1 Digital Testing

The digital SAT Suite represents College Board's full shift to digitally based testing for its flagship college and career readiness assessments. All students—with the important exception of those requiring paper-based accommodations for fair access to the tests—take the SAT Suite tests in Bluebook.

This embrace of digital testing for the SAT Suite offers several critical benefits to those who take the tests, administer the tests, and use the tests' data.

First, the shift recognizes that today's students live much of their lives digitally, including how they learn in school, connect with friends and family, find information to answer their own or assigned questions, and spend their leisure time. The SAT Suite should not be the exception to this. By moving the tests to a digital format, College Board sought to meet students where they are, using a modality (digital delivery) that is increasingly familiar and comfortable to students.

Second, digital testing using the multistage adaptive testing process discussed in detail in Section 4.3.1, Multistage Adaptive Testing, allows College Board to give much shorter versions of the SAT Suite tests relative to their paper and pencil predecessors while maintaining scoring precision (accuracy) and reliability. Students and schools benefit when the same quality of college and career readiness testing that College Board is well known for is compacted into a reduced time frame.

Third, digital testing greatly streamlines the test administration process. Schools and other test centers giving the digital SAT Suite tests no longer have to receive, sort, securely store, re-collect, and ship back test booklets. Thanks to College Board's Test Day Toolkit app, test proctors' work has also been significantly simplified. Notably, the critical function of test timing has been turned over to Bluebook, ensuring that all test takers have exactly the same amount of time to test and are able to track precisely the time they have left via a built-in timer (which can, at students' discretion, be hidden until the five-minute mark in each test module).

Fourth, the shift to digital testing facilitates the expansion of test administrations and the enlargement of testing windows as part of school day testing. The innovations implemented by the digital SAT Suite will allow for more testing opportunities for individual students as well as much greater flexibility on the part of state and district SAT Suite users in scheduling testing to fit the needs of their schools. Critically, this expansion of testing opportunities does not come at the cost of test security, as the digital SAT Suite administers highly comparable but unique test forms to every student.

Fifth, the move to digital testing will, soon after the initial operational administrations, enable faster score reporting than was possible with paper-based testing. Users will get the data they need to inform decisions much sooner than with the paper and pencil SAT Suite.

## 4.2.2 Bluebook

College Board administers the digital SAT Suite on Bluebook, a proprietary digital testing application customized for the digital SAT Suite. This app is a modified version of the one used to successfully deliver AP Exams. Having a customized and well-vetted test delivery application allows College Board to fully meet SAT Suite users' needs and to respond in an agile manner by quickly making updates and refinements as needed.

Bluebook has been designed to conform to the principles of Universal Design for Assessment, or UDA (Thompson et al., 2002). UDA is a set of principles grounded in prior work on universal design for accessibility in other fields—notably architecture, where the concept originated. The overriding goal of UDA is to purposefully make tests as accessible as possible to the largest number of people so that the maximum number of test takers have full, unimpeded access to the tests and their content. Where the application of universal design principles and the offering of universal test-taking tools is insufficient to allow some test takers that level of access, accommodations and supports are provided to level the playing field in ways that welcome students into the tests without compromising the constructs (concepts) the tests are designed to measure, simplifying the assessment tasks, or inadvertently providing the answers to items.

The tests of the digital SAT Suite meet the following requirements of UDA:

1. **Inclusive assessment population.** The digital SAT Suite offers "opportunities for the participation of all students, no matter their cognitive abilities, cultural backgrounds, or linguistic backgrounds" (Thompson et al., 2002, p. 6) by making testing device accessibility features and universal tools available in Bluebook for all students; offering a wide range of accommodations and supports to students who require them, including members from special-needs populations (including English learners and students with disabilities) in pretesting and other studies; and engaging directly with special-needs populations via studies targeted at better understanding their requirements and preferences.
2. **Precisely defined test constructs.** The digital SAT Suite carefully articulates constructs to promote fairness and accessibility by differentiating between the skills and knowledge that are appropriate to assess (i.e., *construct relevant*) and confounding elements that may impair an accurate assessment of those skills and knowledge (i.e., *construct-irrelevant factors*).
3. **Accessible, nonbiased test items.** The digital SAT Suite endorses the notions that test makers should "[incorporate] accessibility as a primary dimension of test specifications" (Thompson et al., 2002, p. 9, citing Kopriva, 2000) and verify that test questions are free from bias (i.e., construct-irrelevant factors that may influence test performance in unintended ways).
4. **Amenable to accommodations.** The digital SAT Suite offers a wide range of accommodations and supports for students who require them.

5. **Simple, clear, intuitive instructions and procedures.** The digital SAT offers test instructions and procedures that are "easy to understand, regardless of a student's experience, knowledge, language skills, or current concentration level" (Thompson et al., 2002, p. 13), provides students with practice opportunities prior to testing, and makes sure that test administration conditions are well documented so that they can be standardized and consistently replicated.
6. **Maximum readability and comprehensibility.** The digital SAT provides texts that are no more linguistically complex than they need to be to satisfy the demands of the construct being measured and features items that "use plain language when vocabulary level is not part of the construct being tested" (Thompson et al., 2002, p. 15).
7. **Maximum legibility.** Digital SAT Suite text and graphics are clear and legible, and item response formats were designed with the needs of all test takers, including those with visual impairments or issues with fine motor skills, in mind. The line reader tool allows the student to focus on one or more lines of text on the screen.

The concept of universal design in assessment and how it applies to the tests of the digital SAT Suite are discussed in greater depth in Chapter 7, Fairness.

## 4.2.3 Accessibility

Accessibility is a critical aspect of test fairness. In addition to the aforementioned universal design principles, the digital SAT Suite advances the goal of maximal accessibility for all students through the provision of universal tools and the availability of accommodations and supports for those students who require them.

### Universal Tools

Bluebook supports a number of universal tools that all students, at their discretion and preference, may use or not use to improve their test-taking experience. These tools include a built-in version of the DESMOS Graphing Calculator, an annotation tool, an answer choice elimination tool, and a method of marking items to be reviewed before time elapses. By design, some of these universal tools, such as those for zooming in and out, were previously offered only as accommodations (e.g., large print); their universal availability in the digital SAT Suite serves to increase the accessibility of the tests for all students. See Chapter 7, Section 7.3.2, Universal Tools, for more information.

### Accommodations and Supports

To ensure that a fair testing environment is available to all test takers, College Board provides students with disabilities taking the digital SAT Suite assessments with the accommodations that they need. This practice ensures that when appropriate and possible, College Board removes or minimizes construct-irrelevant barriers that can interfere with a test taker accurately demonstrating their true standing on a construct (AERA et al., 2014). The digital SAT Suite continues to offer the same range of accommodations and supports previously available in the suite's paper and pencil format, with the caveat that some tools formerly available only as accommodations in paper-based and linear digital testing, such as the ability to zoom in and adjust contrast, are available to all test takers as universal tools, which students may elect to use or not use. Note that students also have access to both adaptive and linear practice tests in Bluebook. The linear practice tests can also be downloaded at no cost from the College Board website. Although these tests are somewhat longer than their adaptive counterparts, they are otherwise built to specifications highly comparable to those used to construct the digital tests, making them effective practice for students who will be taking the linear format due to testing accommodations. The accommodations and supports offered by College Board are discussed more fully in Chapter 7, Section 7.3.3, Accommodations and Supports.

### Practice

Another benefit of Bluebook is that students are able to take full-length adaptive practice tests for free in the same application in which the vast majority of them will take operational SAT Suite tests. This feature enhances the value of practice by allowing students not only to assess the current state of their skills and knowledge but also to gain experience and comfort with the exact way in which they will be assessed on test day. Students with accommodations can also practice with these accessibility features enabled on their testing devices and select extended time and breaks in accordance with their approved accommodations. In addition, the small number of students who will take digital SAT Suite tests in a linear (nonadaptive) format have access to linear practice forms, which provide them with practice opportunities identical to what they will encounter on the actual test.

# 4.3 Form Assignment and Administration

## 4.3.1 Multistage Adaptive Testing

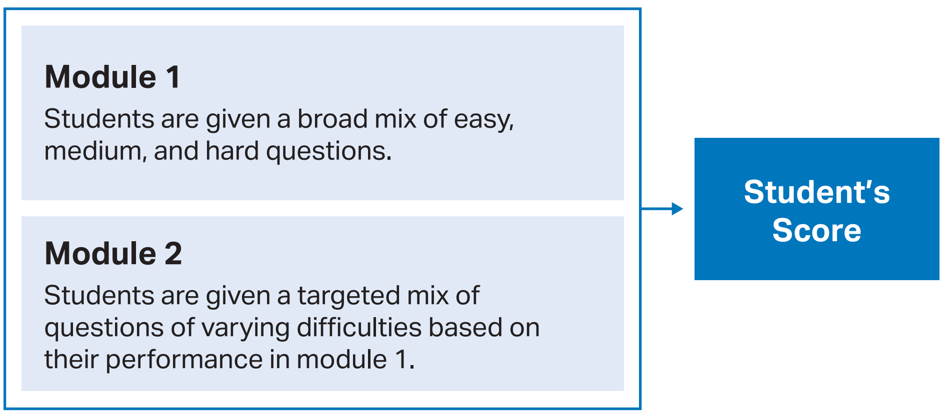
For the digital SAT Suite, College Board has shifted from a linear testing model as the primary mode of administration to an adaptive one. In linear testing—the traditional approach for the SAT Suite—a student is given a test form whose array of items has been set prior to test day and does not change based on the student's performance during the test. Linear tests of this sort are attractive to test makers in part because they allow for fine control of the content presented to students, but they have notable drawbacks as well. In addition to being vulnerable from a test security standpoint, linear test forms are fairly lengthy. This is because the test developers devising such a form cannot make any assumptions about students' achievement levels and therefore must include items across the full range of the test's specified difficulty to measure accurately any one student's achievement. A linear test form is, in a real sense, a "one-size-fits-all" testing model—functional, time tested, but inefficient both at scale and for individual students.

In an adaptive test model, by contrast, the digital testing application adjusts the difficulty of the items given to students based on the performance of individual test takers. These adjustments help ensure that any given student on test day is being administered items of difficulty levels appropriate to their level of achievement. In contrast to linear testing, in which students (particularly those at the high and low ends of the score distribution) often end up being given items that are either too easy or too hard for them, adaptive testing adjusts item delivery according to what the application "learns" about a given student during the exam. Because of this more precise targeting to student achievement level, adaptive testing allows for shorter tests than their linear counterparts and, critically, does so without loss of measurement precision or test reliability. Adaptive testing for the digital SAT Suite is a win for students (and many others) because it means shorter tests that are just as accurate as longer ones.

The digital SAT Suite uses a multistage adaptive testing (MST) model. An MST model segments the testing experience into distinct *stages*, with each stage composed of a *module* of test items. The first module in an MST test such as those of the digital SAT Suite typically consists of items across a broad span of difficulty (i.e., easy, medium-difficulty, and hard items) so that a robust if provisional assessment of test taker achievement can be obtained. The app then uses this information to select the next module to administer. This module consists of items targeted to the test taker's performance up to that point by being either, on average, more or less difficult than the items in the first module. This process continues throughout a given test's stages until a final measure of the student's achievement is obtained. The set of a given initial-stage module and its associated subsequent-stage modules is known as a *panel*.

For the digital SAT Suite, College Board employs a simple two-stage MST model, depicted schematically in Figure 4.1.

**FIGURE 4.1** DIGITAL SAT SUITE MST MODEL




On all digital SAT Suite tests, students begin each test section (Reading and Writing; Math) by answering items in the first module of a given panel. This initial (routing) module comprises half the items of each test section and consists of a broad mix of easy, medium-difficulty, and hard items. These items are sufficiently numerous and diverse to obtain an accurate if provisional measure of a given student's achievement level on the test section. Items from all four Reading and Writing content domains are included in both modules of the Reading and Writing section, and items from all four Math content domains are included in both modules of the Math section. This helps ensure that students are sampled fairly on all key content dimensions in the first module prior to being routed to the second in each section.

Based on student performance on the initial module, Bluebook selects one of two potential second-stage modules to administer. One such module consists of a targeted mix of items that are, on average, of higher difficulty than those in the first module, while the other consists of a targeted mix of items that are, on average, of lower difficulty than those in the first module, although both options include easy, medium-difficulty, and hard items, albeit in differing proportions. Once students have answered the items in the second module, testing on that section is complete, and a section score can be calculated based on student performance across all the items given in the section.

MST testing benefits students in two main ways. First, it results in shorter tests that retain the precision and reliability of longer (linear) tests. Second, students taking one of the digital SAT Suite tests can navigate freely through a given module's items, previewing upcoming items or marking earlier items to return to should time permit.

## 4.3.2 Embedded Pretesting

The digital SAT Suite incorporates embedded pretesting into its design. In embedded pretesting, a small number of pretest (unscored) items are included, or *embedded*, among the operational (scored) items. Although they are not administered for a score, these pretest items are otherwise indistinguishable to students from the operational items on which their scores are based. This ensures that students give maximum attention and effort to these items, which enhances the predictive power of the pretest statistics yielded. The number of pretest items in each test form is kept intentionally low so that students are not unduly burdened with answering items that do not contribute to their score but is still high enough that College Board can continue to offer students the same high-quality digital SAT Suite testing experience indefinitely.

## 4.3.3 Discrete Items

All items on the digital SAT Suite are in a discrete (standalone) format, meaning that students are able to answer each item independently, without reference to a common stimulus such as an extended passage. This represents a departure from the paper and pencil SAT Suite, which used a combination of discrete items and item sets.

The decision to use discrete items exclusively on the digital SAT Suite was prompted partly by the nature of College Board's digital testing model. An adaptive test model, such as the multistage model employed for the digital suite, operates more efficiently when choices about what test content to deliver are made in small rather than larger units. Moreover, these small units can be flexibly combined to create large numbers of highly comparable but nonetheless unique test forms, thereby enhancing test security.

At the same time, the shift to exclusively using discrete items offers several key benefits for students and for the assessments themselves, particularly with respect to the digital SAT Suite Reading and Writing section. First, the shift reduces the amount of cognitively low-level skimming and scanning required to answer reading and writing test items, since all the information needed to answer each item is contained within a brief passage or pair of passages. Test takers can instead focus on demonstrating higher-order reading, writing, analysis, and reasoning skills, such as inferring, rather than spending time searching for relevant information in a longer passage. Second, students who might have struggled to connect with the subject matter of a long passage and then answer up to 11 items about it in the Reading and Writing section of the paper and pencil tests can, on the digital tests, simply give their best answer to each item and move on, knowing that one wrong answer will not negatively impact their scores materially. Conversely, the inclusion of more passages, and therefore more topics, dramatically increases the likelihood that students will find subjects of interest to them on the tests, which will keep them more engaged during testing. Third, the use of discrete items eliminates the possibility, however remote, that items within a set linked to a common stimulus may interact with one another, such as one item inadvertently cluing the answer to another item. Finally, the use of discrete reading and writing items linked in passage-dependent ways to brief stimulus texts obviates the value of certain test preparation strategies intended to short-circuit the intended rigor of tasks, such as not closely reading the entire stimulus and instead attending only to those portions directly relevant to answering particular items.

Importantly, the shift to discrete items has not entailed a reduction of test rigor. Though shorter, passages on the Reading and Writing section of the digital SAT Suite tests are still selected to represent the same range of text complexities correlated with college and career readiness requirements as in the paper and pencil tests, and they continue to sample from and represent the norms and conventions of texts produced in a wide range of academic subject areas, including literature, history/social studies, the humanities, and science. Moreover, pretesting of digital SAT Suite items in both Reading and Writing and Math has consistently demonstrated that digital-suite and paper-suite items are of highly comparable difficulty, and both suites' tests emphasize higher-order thinking skills over low-level recognition, recall, and rote application of rules and algorithms. Furthermore, in 2023 College Board undertook a rigorous qualitative study of students' thinking processes while taking the tests (College Board, 2024), which confirmed that, as was the case with the paper-based SAT (College Board & HumRRO, 2020), the digital SAT Suite tests elicit the kinds of higher-order reading, writing, and math skills and knowledge required for college and career readiness. For more details on this study, see Chapter 8, Section 8.3.1, Cognitive Labs.

# 4.4 Security Protocols and Analyses

## 4.4.1 Test Security Rationale

Although the digital SAT Suite can open doors for students and connect them to opportunities they might otherwise miss, it can only do so if the tests themselves are secure and the results are accurate reflections of students' own efforts. Test security challenges, which are infrequent but highly consequential, threaten the integrity of the tests and the confidence that test takers and data users have in them. Over the long term, these threats, if unmet, erode trust in the tests. In the nearer term, they risk curtailing students' access to testing, as they can have a potential impact on scoring, up to and including rare situations in which scores or whole administrations are canceled due to security compromises.

A key motivation behind College Board's introduction of the digital SAT Suite was to meet these security challenges head-on and to do so in a way that actually expanded, rather than restricted, access to the tests. The digital-suite tests reduce test security risks in a number of important ways, notably by eliminating the need to physically deliver, handle, store, distribute, collect, and reship paper test materials around the world and by ensuring that each student who takes one of the digital tests is administered a highly comparable but unique version of the test.

## 4.4.2 Security Features

### More Secure Test Design

The tests of the digital SAT Suite are more secure than the paper and pencil tests they have replaced. As mentioned above, the switch to digital has eliminated the paper handling that not only places burdens on test administrators but also creates security risks. Bluebook also displays only one test item at a time, making it much more difficult for bad actors to surreptitiously photograph or otherwise copy test content. Most critically, though, the digital SAT Suite assessments have been designed and developed such that each student is administered a highly comparable but unique version of the test. Additionally, there is a greater variety in the items that are presented to test takers. These innovations greatly diminish any value in students copying from their test-taking neighbors or scouring the internet for leaked test content.

## Test Administration Security Procedures

Training for test coordinators who administer the test is required and managed through College Board systems. On the day of testing, test takers will be expected to provide a valid photo identification in addition to their admission ticket for admittance into the test center. Students will not be granted entry without the appropriate security documentation. During test administration, proctors will actively monitor students to ensure that all security procedures are being followed. Failure to comply with security procedures may result in a test taker's dismissal from the test center and/or their scores being withheld or canceled. Up-to-date information about test day protocols and procedures can be found on College Board's digital SAT Suite website, **satsuite.collegeboard.org/digital**.

## Additional Security Measures

College Board executes several additional procedures to ensure test security. These procedures include test center audits, policies that prohibit students from accessing programs and applications other than Bluebook, and posttest analysis. Details of these procedures are kept confidential to maintain their efficacy as security measures.

Chapter 5

# Test Scoring and Reporting

## 5.0 Introduction

This chapter covers the scoring procedures for the digital SAT Suite assessments and the means by which those scores are reported. Section 5.1, Digital SAT Score Procedures, covers the manner in which the digital-suite tests are scored and how scores are generated in a way that is reliable and fair to all students. Section 5.2, Digital SAT Score Reporting and Resources, discusses the reporting of those scores and the various resources available to students and educators to help them understand, interpret, and make use of those scores.

## 5.1 Digital SAT Score Procedures

### 5.1.1 Scoring Process

#### Student Response Entry

Test items on the digital SAT Suite require students[7] to provide their response in one of three ways in Bluebook. The first, multiple-choice (MC) items, requires students to select their answer from one of four given choices. They may select their answer using a mouse, touchscreen, keypad, stylus (iPad, Chromebook, or Windows device), keyboard shortcuts, trackpad, or other approved technology. The second, student-produced response (SPR) items, require students to enter their answer into a designated field that follows the item prompt. The student is presented with a preview of their typed response so that they can verify that the answer they entered was what they had intended to enter. Directions for SPR items appear alongside each item that requires a student-produced response. Lastly, for the digital SAT Essay, which is administered as part of select U.S. school day administrations, students type their response into the designated text box that follows the prompt.

Students who require paper testing can record or mark their responses in the test booklet or dictate their answers when that accommodation is necessary. (See Chapter 7, Section 7.3.3, Accommodations and Supports, for more information on accommodations offered during digital SAT Suite testing.) Testing center staff are expected to transcribe student answers to MC and SPR items into Bluebook within 24 hours of when the student completed testing; these responses are then submitted for scoring as described below. Students taking the digital SAT Essay on paper write their response on an Essay answer sheet, which is then returned to College Board to be scanned and scored following the process outlined below.

---

7 The following information pertains to students taking the test without accommodations.

Bluebook is designed to prevent loss of a student's work and be an efficient use of testing time. If a student loses Wi-Fi connectivity during testing, their answers will be encrypted and saved to the device they tested on, with no loss of testing time. If they do not have a connection when time runs out, they will be given instructions to sign into Bluebook when they are back online and to click/tap the Submit Answers button on their home page.

College Board has built a resilient system that is capable of responding and recovering from intermittent network interruption once the test has begun.

## Production of Test Scores

College Board has quality-control processes in place to ensure the accuracy of scoring of MC and SPR items. Before test packages are made available for students to download, manual as well as automated quality control checks are performed on all the packages to ensure that item metadata, including scoring keys, is correct. Once students take the test and submit packages, the primary scoring engine scores these packages and generates item scores (1/0), ability scores, scale scores, and other scores (such as domain). Automated end-to-end quality-control procedures, which include item level (1/0 for both MC and SPRs) and scaled scores, is performed on 100% of the packages. A parallel scoring quality-control system receives result packages from Bluebook and performs item scoring and generates scale scores using a second independently developed and separately hosted Quality Control (QC) system that scores the item responses and generates the scaled scores. The QC system compares its output with that generated from the scoring system. Any discrepancies in the item, ability, or scale scores are flagged for further analysis and resolution. Result packages that pass scoring quality control are further processed to make the scores official and available for reporting.

College Board's scoring system evaluates student responses for each MC and SPR test item in the following ways:

- For MC items, the system compares the student-selected response to the encrypted answer key, scoring the item as 1 (one) if the student response is correct; otherwise, it is scored as 0 (zero).
- For SPR items, the system "cleanses" the student response (e.g., removes double slashes or decimals) and compares it to the keyed answer(s). If a match is found, the system scores the item as 1 (one), and no additional comparisons are performed. If no match is found, the item is scored as 0 (zero).
- Items that are not attempted are scored as 0 (zero).

College Board makes use of a pre-equated item pool and assembles all panels to meet content and statistical requirements. The pre-equating process requires that all items be calibrated and placed on the same item response theory (IRT) metric. Because all items are on this IRT metric, the item parameters for any set of items can be used to estimate a student's IRT score (theta), which is also on this metric. While theoretically any set of items on this metric can be used to produce equated thetas, College Board uses automated test assembly (ATA) software to ensure that all panels produce highly similar expected theta distributions. The assembly process requires panels to meet content and statistical requirements, including test characteristic curves (TCCs) and test information functions (TIFs) before the test is given.

Under the IRT framework used for the digital SAT Suite, scoring is a function of a student's responses (i.e., based on the pattern of right and wrong answers) and the IRT characteristics (i.e., item statistics) of the set of items administered to the student.[8] Therefore, two students may have the same number of correct answers but have different reported scale scores.

For the Math and Reading and Writing scores, which are administered adaptively, two student ability estimates are obtained for each throughout the duration of the test. The first estimate is used for determining the student's second module at the end of the first module and is based on an expected a posteriori (EAP) estimate. The EAP estimate has several advantages at this stage: 1) incorporates prior information such as historical student performance, 2) avoids optimization problems encountered with other IRT estimates, and 3) handles sparse data better, helping to avoid other estimation issues. These advantages make an EAP estimate less computationally intensive, allowing the estimate to be easily computed on student devices without an internet connection. The second IRT estimate for each section is computed after the testing session is complete and the data has been returned to College Board. This second estimate is based on the maximum likelihood estimation (MLE) using a Newton-Raphson method and is the basis for the student's final score.

Apart from a few state-based accommodations that allow for some students to take only the Math section, all digital SAT Suite test takers take both the Reading and Writing and Math sections. If the student does not wish to answer any item or is directed not to answer any items in the Reading and Writing section, the student will not receive a score for the section that is entirely unanswered. The student will then be able to complete the Math section. Provided there are no irregularities during testing that would require a student to retest or that would invalidate the student's scores, and student's assessment scores are certified, and both the Reading and Writing section score and the Math section score are college reportable. Math section scores for students who take only that section are not college reportable.

College Board's QC team directly oversees and manages verifications and validations that ensure the accuracy of digital SAT Suite scoring processes. In addition to designing and executing the quality-control inspections that confirm the success of operational processes performed both in-house and by vendors, this team advises and collaborates with colleagues throughout College Board who perform scoring-related quality-control inspections, thus ensuring that processes upstream and downstream of the team's involvement are conducted with a full understanding of the requirements. In addition, the QC team maintains a register of existing quality-control processes and works across the scoring processes to identify and remediate gaps, resulting in a continuous improvement program that protects the integrity of the scores that College Board delivers to students and institutions.

### 5.1.2 Scoring of the Digital SAT Essay

Student responses to the digital SAT Essay are received as part of the test package results.  Essay response data are then provided securely via a text file to the essay scoring vendor for human scoring. Essay responses are evaluated on three

8 A fuller discussion of the item parameters and models College Board uses can be found in Chapter 6, Section 6.2, Scaling and Norming.

dimensions: Reading (based on the student's demonstrated comprehension of the passage provided in the prompt), Analysis (based on the quality of the student's analysis of that passage), and Writing (based on the writing quality of the student's response). Raters evaluate how well students' responses demonstrate a careful understanding of the passage, effective and selective use of textual evidence to develop and support points, clear organization and expression of ideas, and a command of the conventions of Standard English. Three dimension scores are reported, each on a scale of 2–8, which is the combined scores of two human raters using the three 1–4 scales established in the Essay's scoring rubric. In cases in which a third, more senior rater is required to adjudicate between discrepant scores (i.e., scores diverging by more than a point assigned by the initial two raters on one or more dimensions), the third rater's 1–4 score on each affected dimension is doubled, and this doubled score is assigned as the final score on each such dimension.

College Board partners with a vendor that uses an online distributed reader model for scoring the Essay. Student demographic and personally identifiable information are not accessible to readers in the distributed scoring system. The distributed reader platform does not allow readers to print out Essay responses. College Board maintains student confidentiality throughout the online scoring process unless exceptional circumstances, such as security violations or when we believe in good faith that it's necessary to protect a student's safety or the safety of others, warrant disclosure to scoring management.

## Rangefinder Training Process and Quality Assurance

All Essay rangefinders are College Board staff. All such staff are highly credentialed, holding undergraduate and often graduate degrees in English and/or related fields. All College Board staff involved in rangefinding have extensive experience in direct-writing assessment and scoring. Since College Board staff, in collaboration with expert consultants and writing specialists, created the Essay task and rubric, they are in the best position to ensure the consistent application of scores across Essay prompts.

College Board staff conduct rangefinding sessions for all training papers used in the operational scoring of the digital SAT Essay. This includes creating anchor and practice sets for a baseline prompt that raters must train and become certified on before they are allowed to participate in actual scoring. Rangefinding sessions are also held by College Board staff to identify anchor, practice, validity, and calibration sets for all prompts used in operational scoring. Rangefinding is intended to produce high-quality training materials that illustrate the prevalent response features of each score point for each digital SAT Essay dimension (Reading, Analysis, and Writing). Anchor sets consist of 10 to 12 exemplar responses that illustrate the score points as well as common score combinations. Practice responses allow raters to attempt the successful application of the rubric to each new prompt they encounter. Validity responses are selected by College Board staff to test raters' performance on prescored essays. Calibration responses are used to test fine distinctions between score points and to reaffirm alignment with the digital SAT Essay rubric. College Board staff, as the primary experts in digital SAT Essay scoring, are the exclusive rangefinders for all digital SAT Essay training materials. The rangefinding process includes assembling a collection of field test or operational responses for a particular prompt. Rangefinders then score each response without knowledge of the previously assigned scores, and rangefinding meetings are held by College Board staff to record staff scores and identify the consensus score for each response. Responses with perfect or high levels of consensus are selected for use in operational training.

## Essay Raters

### Rater Qualifications

Essay raters are required to have a bachelor's degree or higher. The majority of such raters are active classroom teachers or have previously taught secondary- or college-level courses that require writing. Raters work remotely online and under the supervision of a scoring supervisor. Scoring supervisors themselves are selected from a diverse pool based on their qualifications and a record of highly successful performance during training, certification, and operational scoring.

### Scoring Supervisor Qualifications

Individuals who qualify above the required passing rates for raters and demonstrate a high level of accuracy during operational scoring may be promoted to supervisor. Scoring supervisors are selected based on their demonstrated ability to accurately apply rubrics (i.e., strong content knowledge and the ability to consistently follow prescribed scoring criteria rather than their own personal preferences), accurately apply the digital SAT Essay multidimensional rubric to student responses, and ability to monitor scoring quality and to interact well with others by providing helpful feedback to improve the overall quality and consistency of scoring.

For both field test and operational scoring, scoring supervisors must meet the following qualification requirements prior to scoring: 60% perfect agreement and 90% perfect-plus-adjacent agreement for each dimension (Reading, Analysis, Writing) on one of two certification sets of Essay responses.

### Scoring Supervisor Responsibilities

Supervisors are expected to score two hours per day for the purpose of monitoring and maintaining their own scoring accuracy. Supervisors need to backread raters based on the vendor's scoring plan. However, once validity statistics provide enough information to evaluate raters' performance, backreading is focused on raters who are scoring outside of the parameters identified in the vendor scoring plan.

# 5.2 Digital SAT Score Reporting and Resources

The digital SAT Suite provides detailed information about student learning by reporting different types of scoring metrics. Each assessment reports two section scores—one for Reading and Writing and one for Math—and a total score that is the arithmetic sum of the two section scores. Students who take the digital SAT Essay as part of select U.S. school day administrations receive three additional scores—Reading, Analysis, and Writing—based on their Essay response. These Essay dimension scores are not combined with each other and do not affect the section and total scores.

Educators can view student scores for the digital SAT, PSAT/NMSQT, PSAT 10, and PSAT 8/9 tests and find record locator IDs for students who need them by signing into the K–12 score reporting portal (**k12reports.collegeboard.org**) and running a Student Roster Report. Students can view the score release schedule (**satsuite.collegeboard.org/sat/scores/k12-educators/score-release-dates**) to find out when their scores will be available. Please note that scores from paper and pencil versions of the SAT Suite, including (where applicable) SAT Essay scores, may still appear on some student score reports. Some students may have scores from SAT Subject Tests taken prior to June 2021, when these tests were discontinued.

When scores for SAT weekend testing are released, students can view their scores online in the student scores portal (**studentscores.collegeboard.org**). Schools also have access to a downloadable score report PDF called "Your Scores" for these students through the online K–12 score reporting portal.

When scores for in-school testing (SAT School Day and in-school PSAT-related assessments) are released, schools will need to let their students know that there are multiple ways to view their scores. As part of school day testing, students will not be contacted by College Board directly unless they are using the BigFuture School mobile app to receive their scores. A downloadable score report PDF called "Your Score Report" is provided to educators for every digital SAT, PSAT/NMSQT, PSAT 10, and PSAT 8/9 test taker. Students can review their PDF score report with educators at their school or by viewing their scores online in the student scores portal (**studentscores.collegeboard.org**) if they have an account. Schools have access to the downloadable score report PDFs for their students through the online K–12 score reporting portal.

Students who took an SAT School Day, PSAT/NMSQT or PSAT 10 assessment and chose to participate in BigFuture School by providing a mobile phone number during testing can also view their scores on the BigFuture School app. For more information, go to **satsuite.org/k12bigfutureschool**. Please note that the BigFuture School app is available to U.S. students aged 13 and older. Students taking the PSAT 8/9 are not eligible. As always, students aged 13 and older may use a personal College Board student account to view their scores as well as additional score insights online. Students and educators can view a sample of the online student scores portal at **studentscores.collegeboard.org/scores/summary**.

College Board sends scores to a student's school, school district, and state department of education, as applicable. College Board sends PSAT/NMSQT scores to the National Merit Scholarship Corporation (NMSC), the test's cosponsor. College Board does not send PSAT-related test scores to colleges, as these scores are not intended to be part of college admission decisions. Students have the option to send both SAT weekend and SAT School Day scores to up to four colleges and higher education institutions for free initially or by placing a for-fee order to send their scores after that window. SAT scores should not be included on student transcripts that will be reproduced and sent to colleges unless the student (if aged 18 or older) or their parent/guardian has granted permission. Students are allowed to withhold scores from college admission and athletic offices, even when colleges ask for them. Please see the College Board Privacy Center (**privacy.collegeboard.org**) for more information.

## 5.2.1 Information About the Official Score Report PDF

### The Official Score Report PDF

In the official score report PDF, the student can view the three scores available from their SAT, PSAT/NMSQT, PSAT 10, or PSAT 8/9 assessment: total score, Reading and Writing section score, and Math section score. On the PSAT/NMSQT, an NMSC Selection Index score is also available. For each score, the official score report PDF includes the average scores of all testers who took that assessment and the All Tester Percentile (formally called the User Percentile) of each score. All Tester Percentiles are based on the actual scores of the past three cohorts of students in their grade who took the same PSAT-related assessment. SAT includes 12th-grade testers only, PSAT/NMSQT and PSAT 10 includes 10th or 11th graders, and PSAT 8/9 includes 8th or 9th graders. The All Tester Percentile ranks for the first digital administrations in 2023–2024 are reported on tests completed anywhere in the world. The PDF score report also includes student knowledge and skill performance in four content domains in Reading and Writing and four content domains in Math and, for U.S. students, insights into careers in their state.

### Vertical Score Scale

All tests in the digital SAT Suite are on the same vertical scale. Being on a vertical scale allows for consistent feedback to be supplied across assessments as a way to help educators and students monitor growth across grades and to identify areas in need of improvement. This level of feedback can help both students and educators engage in the best possible practice for future assessments: strong classroom instruction. Middle schools/junior high schools and high schools can also use this information to evaluate their curriculum. Higher education institutions can use the scores to gain insight into student readiness.

The score scales are as follows:

#### Total Scores

**SAT:** 400–1600

**PSAT/NMSQT and PSAT 10:** 320–1520

**PSAT 8/9:** 240–1440

#### Section Scores (Reading and Writing, Math)

**SAT:** 200–800

**PSAT/NMSQT and PSAT 10:** 160–760

**PSAT 8/9:** 120–720

The score scales are somewhat staggered across testing programs. That is, the scale for each subsequent testing program has a higher “floor” (minimum score) and “ceiling” (maximum score). This feature serves to facilitate vertical scaling by offering students in successively higher grades the opportunity to demonstrate higher levels of achievement.

### Score Ranges

A student’s ability is better represented by score ranges than points. Score ranges are derived from the standard error of measurement and show how much a student’s scores would likely vary if they took a different administration of the test under identical conditions.

## 5.2.2 Information About Knowledge and Skills

Students can view a graphic showing their performance on each of the eight content domains measured on the digital SAT or PSAT-related assessments. There are four content domains in Reading and Writing and four in Math. For each test section, students can view the following:

- The approximate number and proportion of items in each content domain
- A visual indication of how the student performed in each content domain

The latter is based on assigning student performance in each content domain to a performance score band. These bands correspond to those used in College Board's Skills Insight Tool (see below), which allows students to identify areas they may want to focus on in the future to practice and improve.

### Selection Index

Based on PSAT/NMSQT performance (only), the National Merit Scholarship Corporation derives a Selection Index score that serves as an initial screen for students who enter its annual scholarship program. Selection Index scores are calculated by doubling the Reading and Writing section score, adding it to the Math section score, and dividing this sum by 10. Students who take the PSAT 10 or PSAT 8/9 will not receive a Selection Index score or be considered for entry into the National Merit Scholarship Program.

### Career Insights Snapshot

To help all students consider the full range of future options, score reports include Career Insights Snapshot, which lists careers in a student's state that are connected to the student's assessment performance. Each listed career has a bright outlook, pays a living wage in the state, and requires some form and level of postsecondary education. These careers are presented as examples and are neither formal recommendations nor the only career options that students should consider.

### Additional Support

The score report PDF also includes information on next steps the student can take after reviewing their scores. A QR code or link will provide them with more information on tools and services related to these next steps, which include reviewing the additional insights on their scores, Official Digital SAT Prep on Khan Academy, and college and career exploration.

BigFuture School is a free mobile app for students aged 13 and older who take the PSAT/NMSQT, PSAT 10, or SAT School Day in the United States. It is designed to let students get their test scores quickly, receive planning information on colleges and careers, and learn about financial aid and scholarships. If students provided a mobile phone number when they took the PSAT/NMSQT, PSAT 10, or SAT School Day, they can download the BigFuture School mobile app to receive an in-app notification telling them when their scores are available to view.

Students who take the PSAT/NMSQT, PSAT 10, or SAT School Day may also have the opportunity to opt in to Connections, which allows them to get information from nonprofit colleges and scholarship programs that may be a good match. Connections puts privacy first. No personal information is shared with institutions unless a student directly chooses to do so. Individual schools, districts, or states may choose to not provide access to Connections for its students or for students who test at their schools.

Students who choose not to download the BigFuture School app or who do not have a mobile phone can still access their scores and planning information. Educators will continue to download PDF score reports in the K–12 Reporting Portal for all assessments they administer, which can be shared with students. And, as always, students can log in to their personal College Board accounts at **studentscores.collegeboard.org** to get their scores and additional insights about their scores and to explore BigFuture.

## Digital SAT Suite Benchmarks

The digital SAT Suite College and Career Readiness Benchmarks help students and educators assess student progress toward college and career readiness from year to year. Benchmarks help:

- Identify students who are thriving and require greater challenges
- Identify students who require additional academic support
- Inform instructional and curricular enhancements throughout an institution

For more information, see Chapter 8, Section 8.6, Measuring and Monitoring College and Career Readiness with the Digital SAT.

## Skills Insight

Skills Insight provides an easy and intuitive way for teachers, students, and education professionals to better understand and make use of scores from the digital SAT Suite. Skills Insight describes the skills and knowledge that students scoring in particular ranges (performance score bands) on digital SAT Suite assessments are likely able to demonstrate. Test takers can find performance score bands corresponding to their performance on items in the four content domains in the Reading and Writing and Math sections on their score report and then use this information in the SAT Suite of Assessments Skills Insight Tool (**satsuite.collegeboard.org/skills-insight**).

### Skills Insight Components

Skills Insight consists of two main components: skill/knowledge statements and exemplar test items. Skill/knowledge statements describe what students scoring in particular performance score bands on the Reading and Writing and Math sections of the digital SAT Suite tests are typically able to demonstrate. Exemplar test items concretize those statements by illustrating the kinds and rigor of test items that students in various performance score bands are typically able to answer correctly. Examining the statements associated with a student's performance score bands and (if applicable) at higher bands can help students and those working with them better understand test performance and how to improve it.

## Educator Resources

The K–12 Reporting Portal to which educators have access includes two sections: Reports and Downloads. Educators can access one or both sections depending on the K–12 reporting role they have been assigned. To learn more about gaining access to the K–12 Reporting Portal, review the Granting Access and Assigning Roles section of the K–12 Reporting help page (**satsuite.collegeboard.org/help-center/k12-reporting-portal**).

In the Reports section, educators can:

- View aggregate reports on overall student score and benchmark performance, including performance by various demographics
- View reports that provide individual student scores
- Access info on students' usage of BigFuture School in reports
- Print individual PDF student reports or batches of such reports
- Run and export a range of reports at the school, district, and state levels (as appliable)
- Quickly access reports they recently ran
- Filter, sort, and export data

To view scores for individual students as well as aggregate data, educators will need the Detailed Reports role assigned to their College Board professional account by their institution's K–12 Reporting Access Manager. If they do not need access to scores for individual students, they can instead have the Summary Reports role assigned to their College Board professional account.

In the Downloads section, educators can access:

- Raw Data Files for Your Systems, if they have been assigned the File Downloads role by their institution's Access Manager, which has preformatted files they can import into student information systems.
- Reports You Scheduled (Growth Reports or Batch Student PDFs) as well as Custom Reports From College Board (annually delivered cohort reports and other custom reports from College Board), if they have the Summary Reports or Detailed Reports role.

The reports listed below are good choices to aid in student counseling, curriculum and instruction review, or tracking progress.

### Student Counseling

The following reports can be used to design strategies to help individual students develop their skills and knowledge:

- Roster Report
- Student Report

### Curriculum and Instruction Review

These reports help teachers ensure that classroom work aligns to college and career readiness standards:

- Performance by All Students
- Performance by Demographics
- Knowledge and Skills

### Tracking Progress

These reports can be used to develop systems for monitoring the progress of student groups, schools, and districts:

- Performance by All Students
- Performance by Demographics
- Knowledge and Skills
- Growth Report

Educators can find more information about the K–12 Reporting Portal at the Help Center (**satsuite.collegeboard.org/help-center/k12-reporting-portal**).

Chapter 6

# Psychometrics

## 6.0 Introduction

Having previously discussed the process by which reported scores are produced, it is now time look at how those scores are used to establish the numerical systems that convey test performance. Within this chapter, Section 6.1, Adaptive Testing, provides a more technical overview of multistage testing used by the SAT Suite of Assessments. Section 6.2, Scaling and Norming, describes how the reported scale was established and the process to maintain reported normative information. Section 6.3, Reliability and Errors of Measurement, describes the procedures to ensure that adaptive tests produce reliable test scores and consistently route the examinee to an optimal test experience. Section 6.4, Item Analysis, Calibrations, and Pre-Equating, discusses analyzing new field test items before operational use and the pre-equating procedures to ensure comparable scores across operational tests. To that end, Section 6.5, Panel Assembly and Ongoing Psychometric Quality Management, reviews the methods used to monitor the reported score scale and ensure that items used to produce scores remain stable and accurate. The chapter concludes with an overview of test security analytics conducted to ensure valid scores.

## 6.1 Adaptive Testing

The SAT Suite of Assessments uses a form of adaptive testing called multistage testing (MST), a middle ground between traditional linear-based tests and pure item-level adaptive, or computer adaptive (CAT), tests. MST has several components, which are referred to throughout this chapter. While a few of these terms have previously been used in this manual, for the sake of clarity those terms are defined here once more:

**Item Pool:** All test items that have been field tested, analyzed, and approved for use to produce a score.

**Module:** A set of test items grouped together by content and statistical requirements.

**Stage:** A collection of one or more modules of related content; the first stage is called the routing stage and consists of a single module (routing module); later stages often consist of two or more modules that vary by difficulty. A student will take only one module within a given stage.

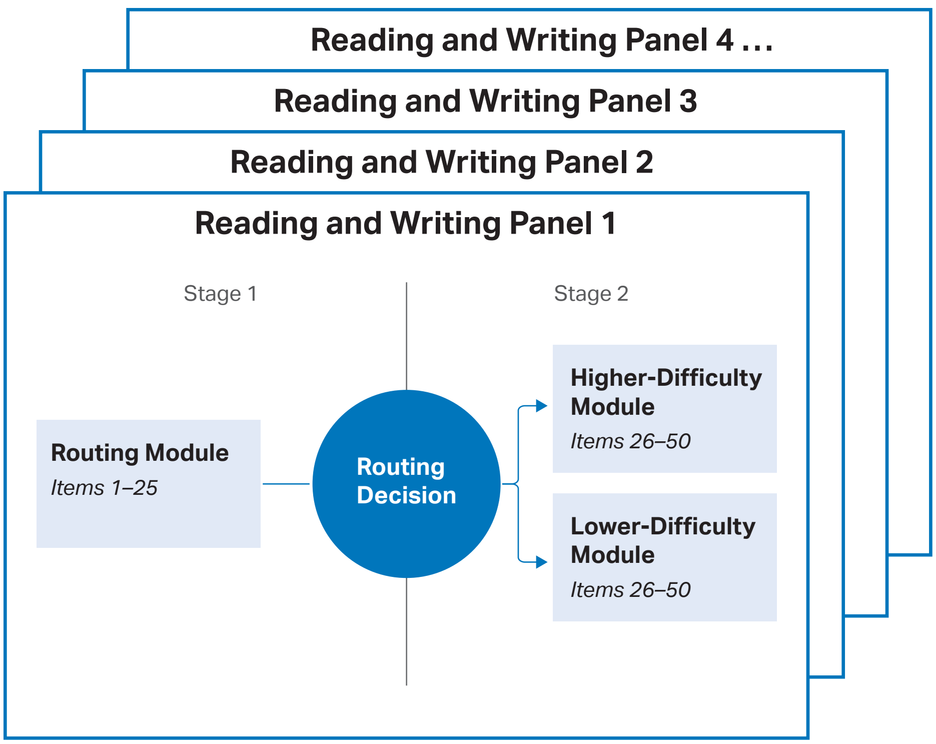
**Panel:** The collection of all stages within a content area that meets all the requirements to measure the construct of interest (i.e., a Reading and Writing panel, a Math panel).

**Routing decision:** The adaptive decision to determine which module in the upcoming stage the student will receive based on the cumulative performance of the student in the previous stage.

**Route:** The path from one stage to the next until the end of the test.

The digital SAT Suite of Assessments utilizes an MST called a 1-2 design (see Figure 6.1). For each section (i.e., Reading and Writing, Math), the student will have a panel that consists of two stages, where stage one consists of a single routing module (consisting of items across a broad span of item difficulty), and stage two consists of a lower- and higher-difficulty module. Depending on the student's performance in the first stage, the student will have one of two possible routes: routing/lower difficulty or routing/higher difficulty, both stage two modules consisting of item difficulty more targeted to the student's performance. To the student, the unique combination of modules a given student sees would be the test or "form" the student was administered. For both assessment sections, each stage consists of 50% of the total items that contribute to the student's score, plus two embedded pretest items in each stage of each section.

**FIGURE 6.1** EXAMPLE READING AND WRITING MST PANELS




MST combines various highly regarded features of the traditional linear and CAT tests. Like traditional linear tests, MST assembles the modules and panels before administration to the examinee. Doing so gives more control over the content that will be provided, ensuring that the content specifications are identical from panel to panel and student to student. MST also allows examinees to review items within a given module, compared to a CAT, where examinees cannot return to previous items. Unlike traditional linear testing, where review can occur across the entire section (i.e., Reading and Writing, Math), MST only allows this within a module. In terms of measurement precision, MST can achieve similar levels of precision with fewer items than traditional linear tests, but typically needs more items than a CAT for comparable levels. MSTs offer other desirable benefits, including better test-security control and fewer computer-processing demands compared to a CAT (Hendrickson, 2007).

As discussed, adaptive testing aims to produce scores with similar or better precision than traditional paper-based tests by using fewer items. In the MST, this goal is accomplished by routing examinees to a module that is optimal in terms of difficulty for the examinee. That is, having examinees take items that are either too difficult or too easy reduces the precision of their scores, in addition to potentially impacting student motivation when the test is overly difficult or too easy.

There are several approaches to making the routing decision in MST. The SAT Suite of Assessments uses a population distribution approach to routing. This approach uses historic archive data of past student performance to determine the median student performance that is then used to set the routing decision point. This approach allows for better exposure control than other approaches. As discussed in Chapter 2, Section 2.3, Statistical Specifications, the statistical constraints used to assemble each panel include maximizing precision at this routing decision point. Section 6.3.1, Simulation Studies, discusses further the accuracy and consistency of the routing decision.

## 6.2 Scaling and Norming

Scaling and norming are used to establish numerical systems that convey test performance and provide context to interpret the scores. The best scales are the ones that support intended interpretations of test performance, which for the SAT Suite of Assessments involves the scale score systems summarized in Chapter 5, Section 5.2 and Section 6.2.2. The most significant part of the scaling work occurred in 2022, when College Board conducted an extensive study with representative students from the United States and internationally. Using the data from the SAT straight-line concordance study of the digital SAT and paper SAT, College Board established a link between the previous paper-based version of the SAT and the digital SAT's item response theory (IRT) metric for the Reading and Writing section score, Math section score, and performance categories in eight content domains. The scores for the PSAT-related assessments are vertically scaled, which is discussed further in Sections 6.2.1 and 6.2.2.

### 6.2.1 Item Response Theory

IRT models were selected to facilitate the use of MST. The previous paper-based version of the SAT Suite of Assessments utilized IRT for form assembly, specifically the three-parameter logistic (3PL) model for select response items and the graded response model (GRM) for the student-produced response (SPR) items in Math. The 3PL model (Lord, 1980) is given by:

$$P_i(\theta_j) = c_i + \frac{1 - c_i}{1 + \exp[-Da_i(\theta_j - b_i)]} \quad (1)$$

where $P_i(\theta_j)$ is the probability of a correct response to item $i$ by a test taker with the ability $\theta_j$; $a_i$ is the item's discrimination parameter or the item's ability to differentiate between examinees; $b_i$ is the difficulty parameter for item $i$; $c_i$ is the lower-asymptote or pseudo-chance score level of item $i$ or the probability that a student lacking complete knowledge would answer the item correctly; and $D$ is a scaling constant that for the SAT Suite of Assessments is set to unity $(D=1)$.

Samejima's (1969, as cited in Thissen and Wainer, 2001) GRM allows for items to be scored in multiple categories; however, when an item is scored dichotomously, the GRM reduces to a special case equivalent to the 2PL model:

$$P_i(\theta_j)=\frac{1}{1+\exp[-a_i(\theta_j-b_i)]} \tag{2}$$

where the parameters in Equation 2 are the same as in Equation 1. The distinction between the special case of the GRM and 2PL is minor, and the GRM is the model used in all estimation and calibration for SPR items within the software utilized.

flexMIRT® (Cai, 2017) is a versatile software package that fits many IRT models using marginal maximum likelihood (MML) to estimate item parameters. The software is also capable of producing several different estimation methods for examinee ability, referred to as theta ( $\theta$ ), including maximum likelihood (MLE) and expected a posteriori (EAP). The EAP method requires additional information about the population of student ability, known as a "prior." Additional information on these methods can be found in flexMIRT documentation (Houts & Cai, 2016).

The IRT models rely on pattern scoring, which incorporates all the parameters of each item administered to the examinee and whether the examinee answered the item correctly or incorrectly. The MLE method is used to determine the student's ability on the complete set of items for a section, which is the  value that maximizes the likelihood function given by:

$$L(\theta_j)=\prod_{i=1}^{nitems}[P_i(\theta_j)]^{u_i}[1-P_i(\theta_j)]^{1-u_i} \tag{3}$$

where  depending on whether the student answered the item correctly (1) or incorrectly (0). The MLE method is used for all examinees regardless of whether they take the MST or a linear assessment version.

For students taking the MST, the routing decision between the first and second stages is determined by the EAP method given by:

$$EAP[\theta]=\frac{\sum_1^q\prod_{i=1}^{nitems}\theta_q\{[P_i(\theta_q)]^{u_i}[1-P_i(\theta_q)]^{1-u_i}\}W_q}{\sum_1^q\prod_{i=1}^{nitems}\{[P_i(\theta_q)]^{u_i}[1-P_i(\theta_q)]^{1-u_i}\}W_q} \tag{4}$$

where $u_i=1\ or\ 0$ depending on whether the student answered the item correctly (1) or incorrectly (0); $\theta_q$ is $q^{th}$ theta value from the prior distribution and $W_q$ is the associated weight for that quadrature point. For all assessments, the prior used is the uniform distribution on a range of [-5,5] with 101 quadrature points. A uniform distribution is considered a non-informative prior, meaning it adds little information about the expected distribution of ability and, therefore $W_q=Constant$. The routing decision, $EAP[\theta]$ on student responses to only the items in the routing module within stage one of each section.

The EAP method is also used to estimate the performance categories for the eight content domains. The prior distribution used for each domain is a normal distribution with a mean and standard deviation. The rationale for a different prior distribution in computing the domain EAPs is due to the number of items contributing to each domain, which ranges from a low of 4 items to as many as 18 across the sections and assessments. For content domains' EAPs, the $W_q$ will be the ordinate of prior distribution at the quadrature point ($q$). Only the items tagged as contributing to the domain are used to compute the associated EAP estimate.

## 6.2.2 Scaling

### Goals for the Scales

The reported section scale scores and content domain performance categories are established as conversions of an IRT theta ($\theta$) to the reported scale score or domain performance categories. The overarching goal for the section scores was to ensure that the scores could continue to be used for the same intended purposes as scores from the previous paper-based version of the SAT Suite of Assessments and have similar meanings.

Reading and Writing (RW) and Math section scores have:

- Ranges of 200–800 for SAT, 160–760 for PSAT/NMSQT and PSAT 10, and 120–720 for PSAT 8/9
- Equal means for the SAT for examinees in the straight-line concordance study
- SAT score distributions that are similar with respect to standard deviations and skewness
- Overall standard errors (SEMs) that are similar to the previous version of the paper-based assessments
- All items answered correctly convert to the highest obtainable scale score for the assessment
- None correct convert to the lowest obtainable scale score for the assessment
- Minimized the gaps in the previous paper-based scale scores

### Scaling of the Digital SAT Suite of Assessments

Three new panels of the digital SAT were constructed and administered in the 2022 SAT straight-line concordance study. After evaluating their properties, one panel was eventually dropped from the study. There were 18,513 nationally and internationally recruited 11th and 12th-grade test takers who participated in the study. Test centers were first recruited to offer the digital SAT. Then, students living within a 50-mile radius of a test center were recruited. Efforts were made to ensure the students recruited represented typical SAT test-takers in terms of ethnic and racial distribution, gender, domestic and international test takers, weekend and school day test takers, and College Board region. To participate, students agreed to take the digital SAT during the study's digital administration times in April and September and take the previous paper-based SAT within one month of their digital SAT.

Three IRT theta estimates were obtained for the Math section and the Reading and Writing section scores, the MLE and EAP estimates previously described, and a test characteristic curve (TCC) estimate. The TCC approach is the sum of the probability of answering every item correctly for a given theta value. This sum can be viewed as the expected true number correct. The TCC theta value is the value associated with the true number correct that equals the examinee's observed number of items correct. For the 3PL, the TCC method has a lower expected true number equal to the sum of the pseudo-chance parameter for all items administered.

Regression and equipercentile methods were used to link the three different IRT theta values to the scale scores on the previous paper-based SAT. The regression method performed well in reproducing individual student scores; however, it resulted in biased estimates and not reproducing the overall distribution of scores, a key goal of the scales. The TCC method produces an unacceptable number of score gaps. The EAP linking produced desirable results; however it was only retained for the routing and domain categories due to having to include information unrelated to student performance in the computation of the IRT theta value. The final linking was completed using the equipercentile method of linking digital IRT MLE theta values to the paper-based scale scores.

The end result of the process was the creation of IRT MLE theta-to-scale score conversion tables for the Math section and Reading and Writing section scores. Because the IRT theta values are continuous on a range of -5 to 5, the theta-to-scale score relationship is a range of theta values that map to the 61 unique scale scores for each assessment. Additional performance category tables were also generated for the Math section and Reading and Writing content domains. The process of establishing the domain performance categories is described later in this chapter. The relationship between theta and scale scores is approximately described as:

$$ScaleScore \cong Intercept + \beta_1\theta + \beta_2\theta^2 + \beta_3\theta^3 + \beta_4\theta^4 + \beta_5\theta^5 + \beta_6\theta^6 + \beta_7\theta^7 + \beta_8\theta^8 \quad (5)$$

Equation 5 is not used to determine the reported scale score; rather it is used to estimate scale score reliability and provide readers with an understanding of the approximate functional relationship between theta and scale scores. The SAT Total score is derived from the Math section scale score (MSS) and Reading and Writing section scale scores (RWSS):

$$Total = RWSS + MSS \quad (6)$$

## Vertical Scale

Vertical scaling allowed the establishment of a common score scale across the SAT Suite of Assessments. The vertical scale across the three assessments of the SAT Suite of Assessments was first established by developing a pre-equated item pool. Pre-equating is described in Section 6.4.2. Three PSAT 10 panels and two PSAT 8/9 panels were assembled from the item pool using content and statistical specifications appropriate for those assessments and used along with the SAT panels from the 2022 SAT straight-line concordance study. These panels were administered to over 26,000 students recruited from grades 9, 10, and 11. The data collection occurred simultaneously with the straight-line concordance study; however, students in the vertical scaling study were not required to take or have previously taken the PSAT 8/9 or PSAT 10.

Examinees were randomly assigned to take one of the three assessments regardless of their grade level, which resulted in all included grade levels taking all three assessments. The common scale assumes that the assessments increase in difficulty as grade level increases, and students in higher grade levels generally have higher abilities. Given the current data collection design, the expected result would be that regardless of the assessment, as grade level increases, the score distribution increases. Within a grade level, it should also essentially be a matter of indifference to which assessment the student takes.

As expected, as grade level increases, one sees an increase in abilities, with increases in the score distributions, consistent in both sections (Figures 6.2 and 6.3). Within grade level, it can also be observed that student performance was similar across assessments (see Figure 6.4 as an example), with slight variations mainly due to differences in content and overall difficulty between the assessment levels.

**FIGURE 6.2** READING AND WRITING SECTION CUMULATIVE THETA DISTRIBUTIONS BY GRADE LEVEL

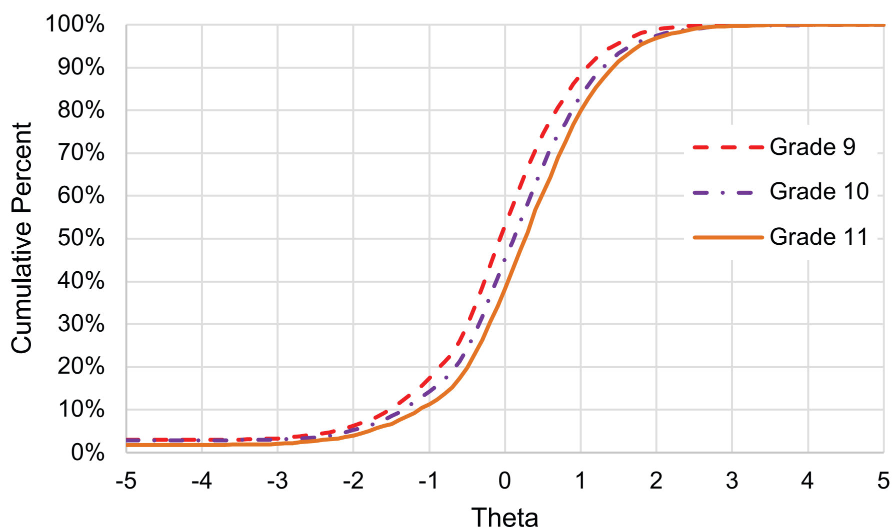



**FIGURE 6.3** MATH SECTION CUMULATIVE THETA DISTRIBUTIONS BY GRADE LEVEL

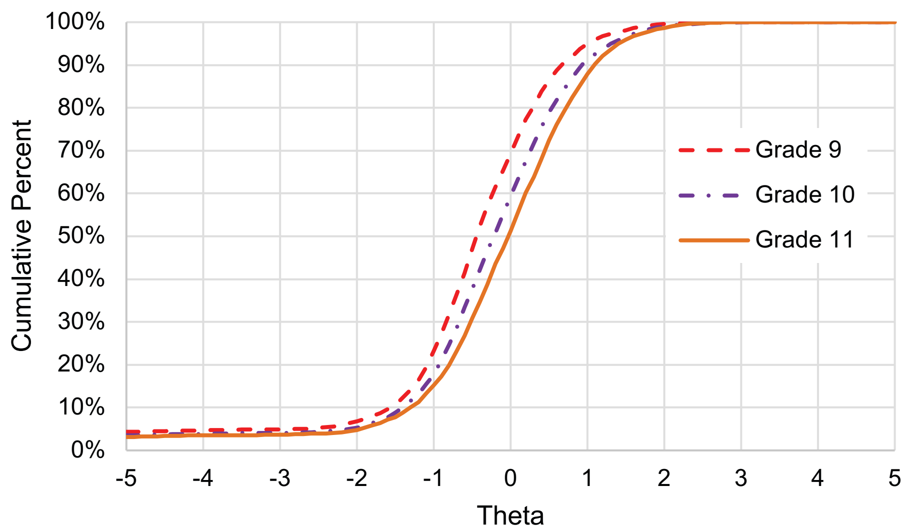
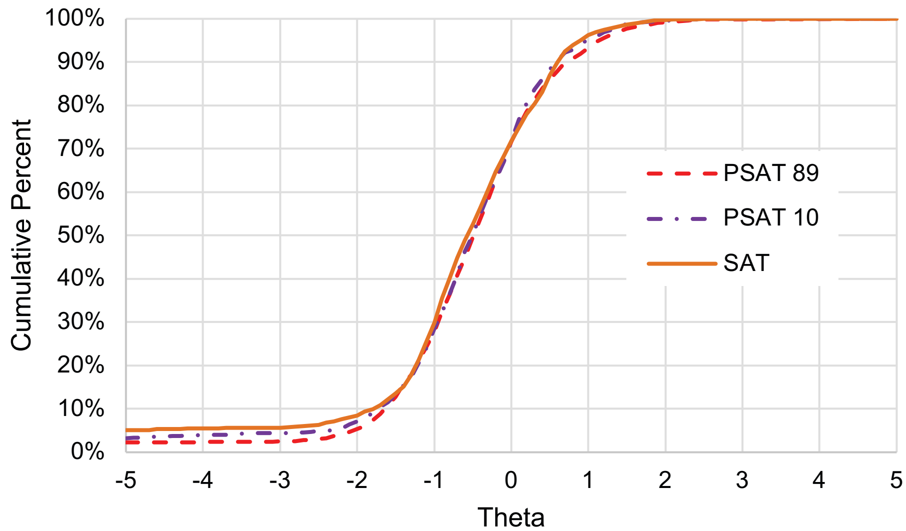


**FIGURE 6.4** GRADE 10 MATH SECTION CUMULATIVE DISTRIBUTION BY ASSESSMENT LEVEL




### Content Domain Performance Categories

The Math section and Reading and Writing section both provide performance feedback to examinees in four content domains. The performance feedback is reported to different stakeholders differently; see Chapter 5, Section 5.2 for more information. Regardless of how the feedback is reported, each domain provides feedback as the student's performance falls within one of seven categories. The categories were determined through a scale anchoring process. The lowest category begins at the scale score associated with the lower 25th percentile of recent PSAT 8/9 score distributions, and the highest category represents the scale score associated with the top 10th percentile of recent SAT score distributions. The intermediate categories are established similarly based on the 25th, 50th, 75th, and 90th percentiles of score distributions of the three assessments.

Because each content domain consists of as few as four items and does not benefit from any adaptive aspects of the test, the precision of the IRT estimates for domains will be low. Users are cautioned about over-interpreting the performance categories. The categories are intended to provide guidance on where students and teachers may focus instructional or practice efforts. Users are also cautioned about making comparisons between students based on performance categories. It is entirely possible, particularly when few items contribute to the domain, that a low-performing student may appear to have a strength in a more challenging domain or perform better than higher-ability students.

## 6.2.3 Normative Information

College Board has made great efforts to link the digital SAT to the previous paper and pencil SAT scale using methods similar to equating. The studies to link these two versions of the SAT assessment involved two studies that recruited 11th and 12th graders, including students who tested as part of school day settings, international students, students with disabilities, and multilingual learners. The study is described in more detail in Section 6.2.2, Scaling. These studies provide strong evidence that the digital SAT scores are distributed similarly to the existing

SAT. Therefore, College Board will continue to use the nationally representative norms produced with the launch of the SAT Suite of Assessments. Please refer to Chapter 6, Section 6.3 of the previous SAT Suite of Assessments Technical Manual (College Board, 2017) for the development of SAT nationally representative norms.

## Norming Updates

To ensure the continued accuracy of norms and score interpretability, and in keeping with Standard 5.11 of the *Standards for Educational and Psychological Testing* (AERA et al., 2014, pp. 104–105), common scale score norms need to be re-established with appropriate frequency. During the initial launch year of the SAT Suite, significant efforts were made to establish normative information that meets all industry and College Board standards. Nationally representative normative information for all three exams—SAT, PSAT/NMSQT and PSAT 10, and PSAT 8/9—will be monitored and updated as needed (as this isn't expected to change rapidly, it will be updated periodically at several-year intervals).

SAT Suite of Assessment user group scale score norms will be updated yearly using a three-year rolling average, dropping the first year when adding new cohort information (i.e., 2023 norms would be based on graduating cohorts of 2021–2023, 2024 norms on 2022–2024 cohorts, and so forth). The user groups updated in 2026 should consist primarily of students who have only taken a version of the digital SAT Suite of Assessments (including linear or paper versions of this assessment). In 2026, College Board will reevaluate the nationally representative norms for the SAT and update as needed.

## Norming Samples

College Board collects information about the following characteristics of the students included in the development and updating of norms:

- Grade level
- Gender
- Race and ethnicity
- First and best language

While College Board collects this information to ensure the norms reflect the desired reference groups, College Board doesn't produce special subgroup normative information, such as for gender, race/ethnicity, or language subgroups. Student characteristics are collected through different processes, depending on how the student is registered for the assessment. Students registering for a weekend administration are asked to complete the non-assessment questions (NAQ), including, but not limited to, the above characteristics. Students who participate in the SAT Suite of Assessments as part of their state/district/school testing program are registered by their school, district, or state. Students may be asked to provide a limited number of demographic characteristics on an optional basis as part of their testing experience to ensure that the characteristics listed in the section above are available to conduct the norming process.

## Norming Inclusion

Inclusion of all test takers, to the greatest extent possible, is a hallmark of the digital SAT Suite of Assessments. With the digital delivery system, the adoption of universal design principles could be implemented more readily. Concerning inclusion of special populations in the norm groups, all students in general—and any student who has tested under a College Board-approved accommodation—will be included in the norm group samples. Indeed, even students who still require a paper and pencil accommodated form are included in the overall population of students for use in all psychometric processes and procedures. The only exception to this is the inclusion of embedded pretest items. Pretest items will only be presented to students who take a digitally delivered multistage version of the assessments.

A few students may take an exam in the SAT Suite with a state-approved accommodation (i.e., an accommodation not approved by College Board). The responses from those students would not be included as part of the overall norm group. The primary reason is that using a non-approved accommodation may alter the construct being assessed.

## 6.3 Reliability and Errors of Measurement

Reliability and errors of measurement are related concepts that measure the consistency in observed scores. Consistency in scores across instances of a test procedure is one aspect that helps ensure that scores are valid for the intended uses. Observed scores can vary for many reasons. This variance can occur, for example, due to a student's state of health, taking a test at two different points in time, taking different forms of the test, or even changes in the test administration. Reliability coefficients are metric-free and range from 0 to 1, with higher values indicating that scores will be very similar across testing conditions if the student's level of knowledge and skills, measured by the test, changes very little. The standard error of measurement (SEM) also quantifies the amount of error within observed scores. However, it is reported on the same metric as the reported scores. The SEM allows stakeholders to understand how much they might expect their scores to change if they retake a parallel test form.

### 6.3.1 Simulation Studies

Simulation studies have an essential role in the development and delivery of adaptive tests. Such studies allow the study of the adaptive algorithm to ensure fairness and evaluate various types of consistency and error likely to be observed in scores. As discussed above, all observed scores contain measurement error; however that error can be decomposed into random error, systematic error, or bias. During the development phase of the testing program, simulations can be used to reduce bias. During ongoing operational activities, simulations can be used to ensure that new panels or linear forms will produce scores with similar levels of error as all other forms, helping ensure fairness.

#### Fairness of Adaptive Multistage Tests

A common concern of examinees and other stakeholders is the fairness of adaptive tests. For MSTs, stakeholders expressed concerns that if students are routed to the lower-difficulty route, they will not have an opportunity to perform well. A simulation study was conducted to address this issue. The crux of the question is, on which set of items would the student have done better if there was no adaptive algorithm?

Exactly 61,000 simulees with known abilities ($SS_T$) were generated, precisely 1,000 simulees for each distinct SAT scale score from 200 to 800. Each simulee had item responses generated consistent with their true ability, or $SS_T$, using the WinGen (Han, 2007) simulation tool. Items from two panels of the Math and Reading and

Writing (RW) sections were used as the true generating item parameters. Each simulee had responses to all items for all modules; therefore, the simulees had 75 RW responses and 60 Math responses. The true item parameters for each panel were used to obtain the final MLE theta estimates for both routes through a panel for both sections. These theta estimates were converted to observed scale scores ($SS_O$) and compared to the $SS_T$, given as:

$$cBias_{Tr} = \frac{1}{I}\sum_{i=1}^{I} SS_{Oi} - SS_{Ti} \quad (7)$$

where $I$ is the number of replications of the simulation, which was 100 in this study. Equation 7 quantifies the amount of bias; and $cBias_{Tr}$ is the conditional bias for $SSTi = T$ ($T = 200, 210, \ldots, 800$), and $r$ is either the lower- or higher-difficulty route. Ideally, bias is close to zero, indicating the score contains no systematic error. Within scale score units, the bias between +/- 5 would, on average, result in the same reported scale score.

Figures 6.5 and 6.6 graph the conditional bias by SST for both routes through the section. Both test figures have a shaded region labeled “Area of Indifference to Examinee,” which is the section of the score scale where either route would produce the same score and the scores would have minimal bias. For higher-ability students, there would be a clear preference to be routed to the lower-difficulty module. For these students, the lower-difficulty module would produce higher (over-estimate) observed scores for both sections but are more biased. For example, a student whose true ability is 680 would expect to score 680 when taking the higher-difficulty route compared to about 700 if they took the lower-difficulty route.

**FIGURE 6.5 CONDITIONAL BIAS BY DIFFICULTY ROUTE: READING AND WRITING**

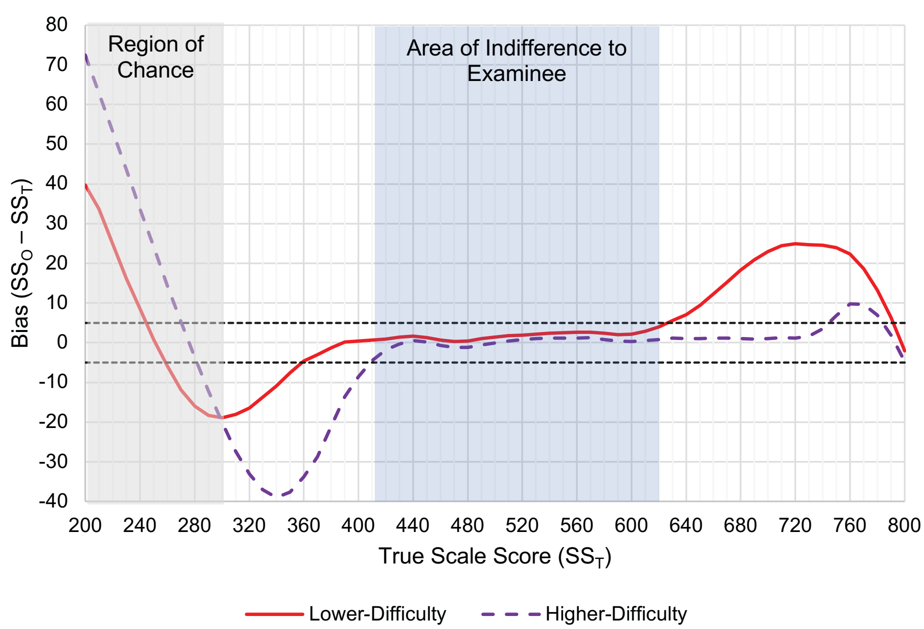
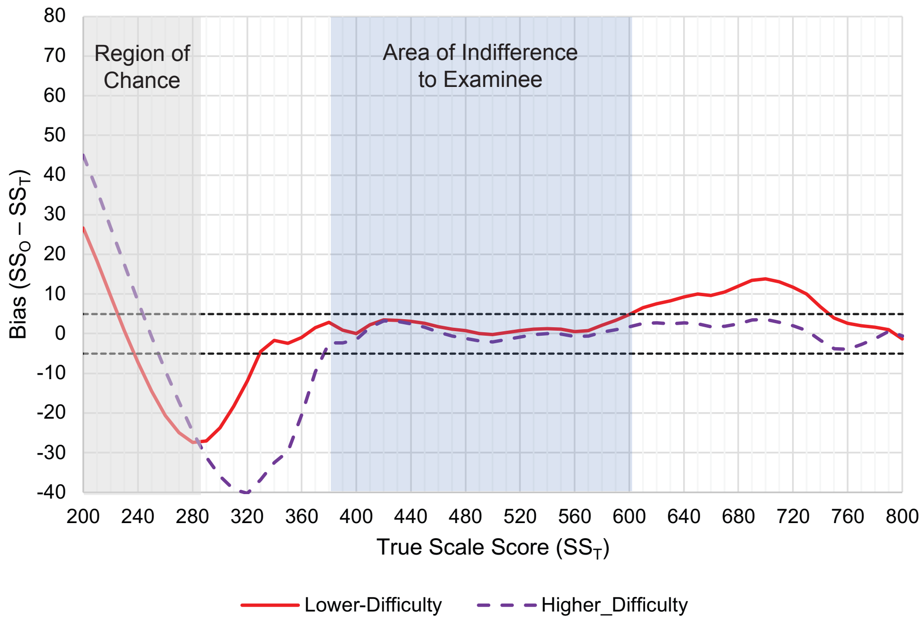


**FIGURE 6.6** CONDITIONAL BIAS BY DIFFICULTY ROUTE: MATH




For students between 300 and 410 in Reading and Writing and between 280 and 380 in Math, there should be a preference by the students to be routed to the lower-difficulty route as the observed scores would also be less under-estimated than by the higher-difficulty route and, in this case, also less biased. For example, a student with a true ability in Math of 320 would expect to earn a 310 by taking the lower-difficulty route compared to a 280 by taking the higher-difficulty route. Both figures also have a shaded region labeled "Region of Chance," which is the region of the scale scores where the random error is the greatest, primarily due to being able to get items correct simply by chance.

The results of the simulation may be counterintuitive in that the simulation provides evidence that students, particularly higher-ability students, would benefit from finding a way to take the lower-difficulty route. However, the simulation results indicate that only higher-ability students would prefer to be routed to the alternative difficulty level in the second stage. To do so, the student would need to answer an unknown number of questions wrong in the routing module of the first stage to be routed to the lower-difficulty module, which, as the simulation shows, even under the best circumstances, would limit the highest possible score they could earn. As such, a student has no reason to attempt to manipulate the system and should always strive to answer as many questions correctly as possible. For its part, the simulation assumes that no routing decision occurred, and that any student would have a choice in which route to take. The simulation was extended to show an example of how such a "choice" might be achieved. The extension imagines that a higher-ability student, knowing that the test is adaptive, will attempt to answer enough items correctly in the routing module to maximize their final score but to be routed to the lower-difficulty module, then perform well on the stage two items. An extreme example is a student who attempts this method and is routed to the lower-difficulty route, then correctly answers all the lower-difficulty items. The results of this experiment indicate that on the Math section, this type of student would, at most, earn a 580, which is within the region of indifference. Students of this ability and higher would likely earn a lower score through this method than if they put forth their best effort. Similar results are seen for the Reading and Writing section.

## Measurement Bias and Error

As previously discussed, all observed scores contain error, quantified either through a reliability index or the standard error of measurement. Simulations allow for the study of the expected error before administering the test to minimize bias and ensure that the standard error measurement is consistent across all panels.

A simulation study was conducted using the same panels used for the simulation described earlier in this section under true administration conditions; specifically, this simulation included the adaptive routing decision. In addition, two linear versions of the SAT were used. The simulation was replicated 100 times, and each simulation generated 10,000 simulees drawn from a normal distribution. For replication and each test section, each simulee had a true theta value, a true route. Two observed final theta estimates were made for the two adaptive panels, and two final theta estimates were made for the two linear panels. By applying the theta-to-scale score relationships, we obtain the corresponding true and observed scale scores and the true and observed College Board's College and Career Readiness (CCR) benchmark classifications.

Similar to Equation 7, the average bias was computed for each replication as:

$$Bias = \frac{1}{N}\sum_{i=1}^{N}\hat{\theta}_i - \theta_i \quad (8)$$

where $\hat{\theta}_i - \theta_i$ is the deviation between the observed/estimated and the true theta value, and $N$ is the number of simulees. In the case of evaluating bias for scale scores, $\theta$ can be replaced with corresponding scale scores in Equations 8.

The standard error of measurement can be approximated in the simulation with the root mean square error (RMSE):

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(\hat{\theta}_i - \theta_i)^2}{N}} \quad (9)$$

Bias and RMSE were computed for the overall test and by route through the test. The overall mean bias was generally small for both sections for both the adaptive and linear versions for both the theta and scale score metrics (Table 6.1). The adaptive and linear versions show meaningful bias in the lower region of the score range, which is related to misestimating the theta values associated with those scale scores (Figures 6.7 and 6.8). This region of scores includes the area of the theta scale where behaviors like guessing or disengagement have an impact on scores.

**TABLE 6.1** BIAS OF ESTIMATED THETA AND SCALE SCORES

| | Reading and Writing | | | | Math | | | |
|---|---|---|---|---|---|---|---|---|
| | Theta | | Scale Score | | Theta | | Scale Score | |
| | Adaptive | Linear | Adaptive | Linear | Adaptive | Linear | Adaptive | Linear |
| Lower Difficulty | -0.06 | | -0.69 | | -0.10 | | -1.18 | |
| Higher Difficulty | 0.06 | | 1.86 | | 0.04 | | 0.78 | |
| Overall | 0.00 | -0.02 | 0.59 | -0.15 | -0.03 | 0.00 | -0.20 | 0.49 |

**FIGURE 6.7** CONDITIONAL SCALE SCORE BIAS: READING AND WRITING

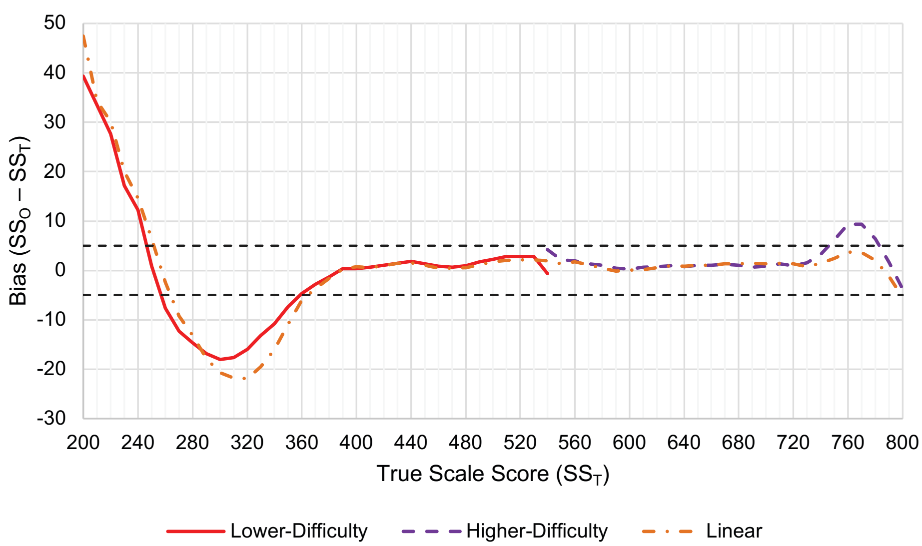
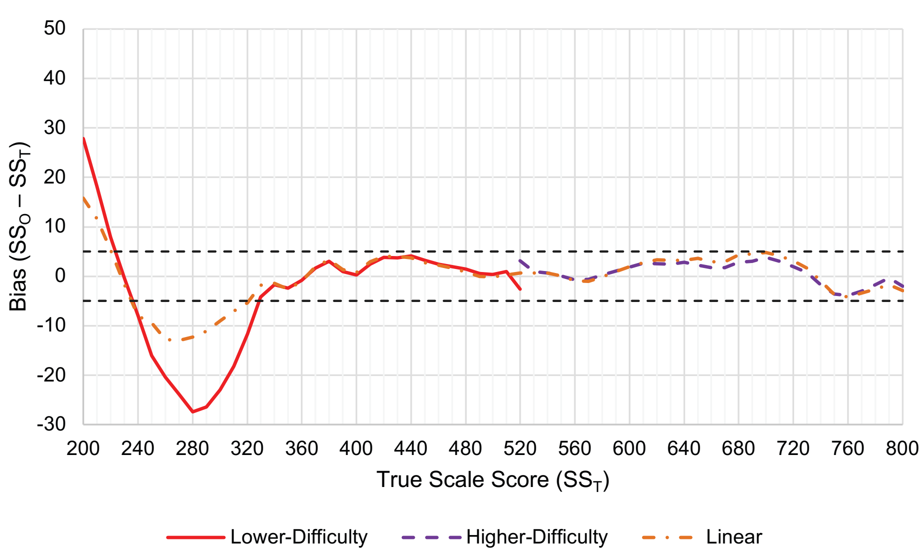



**FIGURE 6.8** CONDITIONAL SCALE SCORE BIAS: MATH



Results of the simulation show that the overall scale score RMSE is similar across the test's adaptive and linear versions (Table 6.2). Within the adaptive test, greater error is present in scores for lower-difficulty routed students, which is consistent with the results of the bias analysis and is most clearly seen in Figures 6.9 and 6.10. The conditional RMSE plots also show that the total error in scores is approximately equal along much of the scale score range.

An essential result of the simulation study is the evidence that the adaptive and linear versions produce scores with similar characteristics. In most cases, it would be a matter of indifference to a student whether they take the adaptive or linear version of the test.

**TABLE 6.2** RMSE OF ESTIMATED THETA AND SCALE SCORES

| | Reading and Writing | | | | Math | | | |
|---|---|---|---|---|---|---|---|---|
| | Theta | | Scale Score | | Theta | | Scale Score | |
| | Adaptive | Linear | Adaptive | Linear | Adaptive | Linear | Adaptive | Linear |
| Lower Difficulty | 0.41 | | 21.13 | | 0.50 | | 21.82 | |
| Higher Difficulty | 0.26 | | 17.66 | | 0.24 | | 19.57 | |
| Overall | 0.34 | 0.35 | 19.47 | 20.56 | 0.39 | 0.29 | 20.73 | 21.03 |

**FIGURE 6.9** CONDITIONAL RMSE: READING AND WRITING

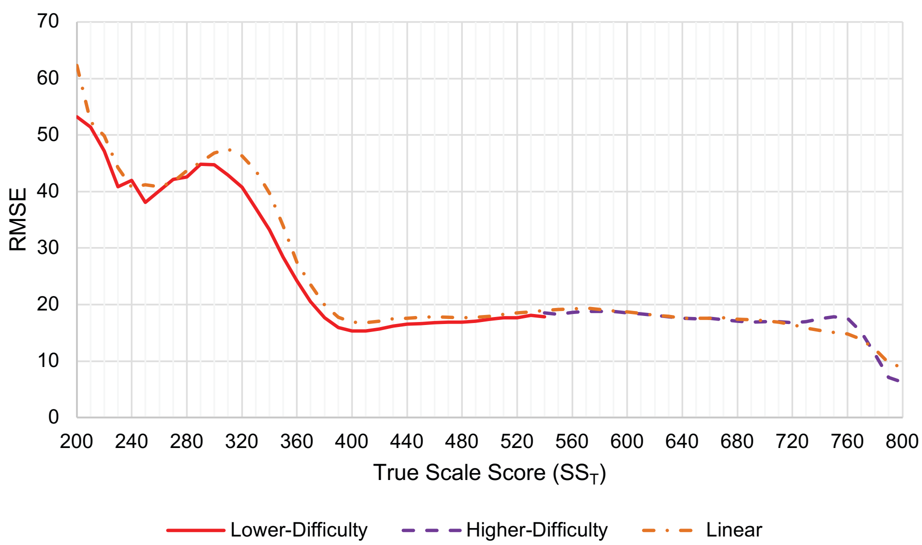
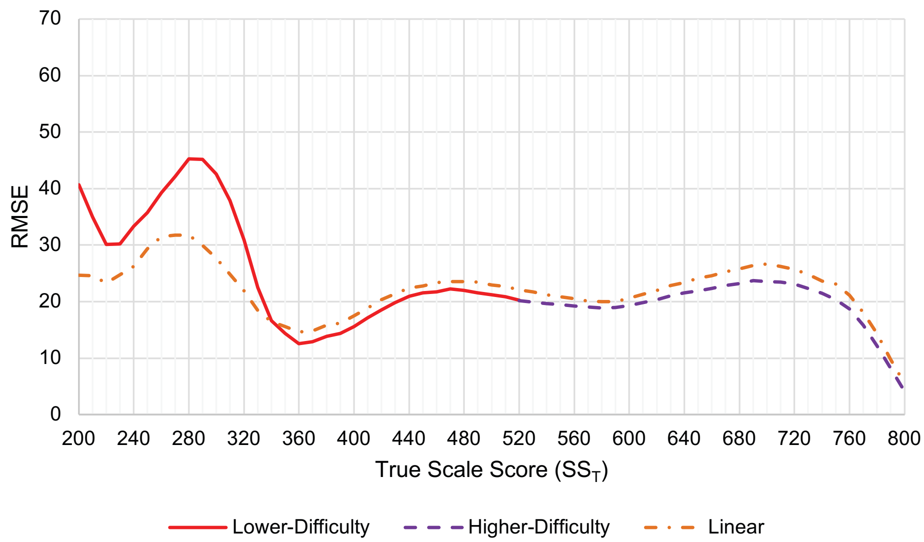


**FIGURE 6.10** CONDITIONAL RMSE: MATH




## Routing Accuracy and Routing Consistency

The simulation study also allows the evaluation of routing accuracy and routing consistency. Routing accuracy (see Table 6.3) is a measure of agreement between an observed route and the true route, or the route had the test scores been perfectly reliable. Routing consistency (see Table 6.4) measures the agreement of two observed routing decisions made on two different test panels. The above-described simulation provides the necessary information to establish the expected routing accuracy and consistency. As seen in the simulation study described earlier in this section, it is important to ensure that the routing decision is accurate so that students are not provided with a test that may unduly under- or overestimate their ability. Routing consistency is essential to have confidence that different panels of the assessment will route examinees in the same way.

Both sections of the assessments have high levels of accuracy, with simulation results indicating that over 93% of all simulees were routed as expected. The Reading and Writing section tended to route more students to the higher-difficulty path than expected. In contrast, the Math section tended to route slightly more students to the lower-difficulty path. Cohen's $\kappa$, which indicates to what degree the agreement is due to chance, is reported for sections. Values closer to 0 indicate chance agreement and values closer to 1 indicate stronger agreement. For both sections, the computed value indicates a non-chance level of agreement. For the Reading and Writing section, about 18% of the total routing inaccuracies occur at the routing point, 82.4% occur within +/- 30 scale score points of the routing decision, and 99.9% occur within the region of indifference (see Figure 6.4). In the Math section, 16.4% of the routing inaccuracies occur at the routing point, 76.6% within +/- 30 scale score points of the routing point, and 99.8% within the region of indifference (see Figure 6.5).

The simulation results also show a high level of agreement for routing consistency, with over 90% of simulees routed to the same path on the two separate panels in both sections. Both sections tended to show slightly higher consistency for simulees routed to the lower-difficulty path. The $\kappa$ values for the routing decision consistency also show strong, non-chance agreement.

**TABLE 6.3** ROUTING DECISION ACCURACY

| | Reading and Writing | | | Math | | |
|---|---|---|---|---|---|---|
| | Observed Route | | | Observed Route | | |
| **True Route** | **Low Diff.** | **High Diff.** | **Row Total** | **Low Diff.** | **High Diff.** | **Row Total** |
| Lower Difficulty | 46.61% | 3.37% | 50.09% | 46.73% | 3.36% | 50.09% |
| Higher Difficulty | 3.43% | 46.59% | 49.91% | 3.39% | 46.52% | 49.91% |
| Column Total | 50.04% | 49.96% | 100.00% | 50.12% | 49.88% | 100.00% |
| Overall Accuracy | 93.20% | 93.25% | | | | |
| Cohen's $\kappa$ | 0.86 | 0.86 | | | | |

**TABLE 6.4** ROUTING DECISION CONSISTENCY

| | Reading and Writing | | | Math | | |
|---|---|---|---|---|---|---|
| | Panel 2 Route | | | Panel 2 Route | | |
| **Panel 1 Route** | **Low Diff.** | **High Diff.** | **Row Total** | **Low Diff.** | **High Diff.** | **Row Total** |
| Lower-Difficulty | 45.23% | 4.77% | 50.00% | 45.32% | 4.77% | 50.09% |
| Higher-Difficulty | 4.84% | 45.16% | 50.00% | 4.82% | 45.09% | 49.91% |
| Column Total | 50.07% | 49.93% | 100.00% | 50.14% | 49.86% | 100.00% |
| Overall Consistency | 90.40% | 90.41% | | | | |
| Cohen's $\kappa$ | 0.81 | 0.81 | | | | |

Routing inaccuracies and inconsistencies will occur, though the simulation results offer strong evidence that these errors will be minimal. When these errors occur, they will occur along the region of the score scale where it should be a matter of indifference to the examinee as to which stage two difficulty level they are routed since the score achieved would be the same (see Figures 6.5 and 6.6).

## College and Career Readiness Benchmark Classification Accuracy and Classification Consistency

College Board's CCR Benchmarks are an essential part of the intended use of the SAT Suite of Assessments. The benchmarks serve as a signal to students if they are on track or likely to be successful in college or have the necessary skills for a career. Furthermore, the benchmarks are often adopted by stakeholders that use the scores as part of their assessment systems or even for federal accountability.

The results of the simulation (described earlier in this section) show that the classification accuracy for the CCR benchmarks in both sections and for both the adaptive and linear versions exceeds 94% (see Tables 6.5 through 6.8). The associated $\kappa$ values are also high. Both versions and sections also tend to classify slightly more simulees at and above the CCR benchmark than expected. The CCR classification consistency is also above 92% for both sections and versions, again with high $\kappa$ values. The adaptive tests tend to show slightly more accuracy and consistency than the linear forms.

**TABLE 6.5 CCR BENCHMARK CLASSIFICATION ACCURACY: READING AND WRITING**

| | Adaptive MST | | | Linear | | |
|---|---|---|---|---|---|---|
| | Observed CCR BM | | | Observed CCR BM | | |
| True CCR BM | Below | At and Above | Row Total | Below. | At and Above | Row Total |
| Below | 30.24% | 2.31% | 32.55% | 30.09% | 2.46% | 32.55% |
| At and Above | 2.23% | 65.21% | 67.45% | 2.41% | 65.04% | 67.45% |
| Column Total | 32.48% | 67.52% | 100.00% | 32.50% | 67.50% | 100.00% |
| Overall Accuracy | 95.45% | 95.12% | | | | |
| Cohen's $\kappa$ | 0.90 | 0.89 | | | | |

**TABLE 6.6 CCR BENCHMARK CLASSIFICATION ACCURACY: MATH**

| | Adaptive MST | | | Linear | | |
|---|---|---|---|---|---|---|
| | Observed CCR BM | | | Observed CCR BM | | |
| True Route | Below | At and Above | Row Total | Below | At and Above | Row Total |
| Below | 48.79% | 2.84% | 51.63% | 48.64% | 2.99% | 51.63% |
| At and Above | 2.50% | 45.87% | 48.37% | 2.68% | 45.70% | 48.37% |
| Column Total | 51.29% | 48.71% | 100.00% | 51.31% | 48.69% | 100.00% |
| Overall Accuracy | 94.66% | 94.33% | | | | |
| Cohen's $\kappa$ | 0.89 | 0.89 | | | | |

**TABLE 6.7 CCR BENCHMARK CLASSIFICATION CONSISTENCY: READING AND WRITING**

| | Adaptive MST | | | Linear | | |
|---|---|---|---|---|---|---|
| | Panel 2 CCR BM | | | Form 2 CCR BM | | |
| Panel/Form 1 CCR BM | Below | At and Above | Row Total | Below | At and Above | Row Total |
| Below | 29.24% | 3.23% | 32.47% | 29.05% | 3.44% | 32.49% |
| At and Above | 3.23% | 64.29% | 67.53% | 3.46% | 64.05% | 67.51% |
| Column Total | 32.48% | 67.52% | 100.00% | 32.51% | 67.49% | 100.00% |
| Overall Consistency | 93.53% | 93.10% | | | | |
| Cohen's $\kappa$ | 0.85 | 0.84 | | | | |

**TABLE 6.8** CCR BENCHMARK CLASSIFICATION CONSISTENCY: MATH

| | Adaptive MST | | | Linear | | |
|---|---|---|---|---|---|---|
| | Panel 2 CCR BM | | | Form 2 CCR BM | | |
| Panel/Form 1 CCR BM | Below | At and Above | Row Total | Below | At and Above | Row Total |
| Below | 47.53% | 3.73% | 51.25% | 47.32% | 3.97% | 51.29% |
| At and Above | 3.81% | 44.94% | 48.75% | 4.02% | 44.69% | 48.71% |
| Column Total | 51.33% | 48.67% | 100.00% | 51.34% | 48.66% | 100.00% |
| Overall Consistency | 92.47% | 92.01% | | | | |
| Cohen's κ | 0.85 | 0.84 | | | | |

## 6.3.2 Standard Errors of Measurement

The standard error of measurement (SEM) provides an estimate of the amount of error or inconsistency in observed test scores. The related conditional SEM (cSEM) estimates the amount of error or inconsistency in observed test scores for a particular true score. A two-step process is used to obtain these measures in terms of reported scale scores. First, an estimate of the IRT cSEM for each examinee is needed for the final IRT theta estimates computed. Then the IRT cSEM is transformed to the scale score using Equation 5. The following subsections provide additional detail.

### Theta

For every test item, it is possible to obtain an error estimate, known as information, and the sum of information across the set of items administered to a student is known as the test information function (TIF):

$$TIF(\theta_J) = \sum_{i=1}^{allitems} \frac{D^2 a^2 [1-P_i(\theta_J)][P_i(\theta_J)-c_i]^2}{P_i(\theta_J)(1-c_i)^2} \quad (10)$$

Where the parameters of Equation 10 are defined the same as in Equation 1, and $\hat{\theta}_j$ is the observed theta for the $j^{th}$ examinee.

The cSEM for $\hat{\theta}_j$ is:

$$cSEM(\hat{\theta}_j) = \frac{1}{\sqrt{TIF(\hat{\theta}_j)}} \quad (11)$$

### Section Scale Scores

Once Equation 11 has been computed, a well-documented property of information is that it is invariant to monotonic transformations (Lord, 1980). Therefore, the test information for the scale score of the $j^{th}$ examinee ($SS_j$) is:

$$TIF(SS_j)=\frac{TIF(\hat{\theta}_j)}{\left(\frac{dSS}{d\theta}\right)^2} \tag{12}$$

where $\frac{dSS}{d\theta}$ can be derived from Equation 5 as:

$$\frac{dSS}{d\theta}=\beta_1+2\beta_2\theta+3\beta_3\theta^2+4\beta_4\theta^3+5\beta_5\theta^4+6\beta_6\theta^5+7\beta_7\theta^6+8\beta_8\theta^7 \tag{13}$$

Equation 13 is simply the slope of the relationship of theta–to–scale scores.

Finally, there is the cSEM for the scale score of the $j^{th}$ examinee:

$$cSEM(SS_j)=cSEM(\theta_j)\left(\frac{dSS}{d\theta}\right) \tag{14}$$

It should be noted that the upper bound of $cSEM(\hat{\theta}_j)$ is set to 1.75, and any value greater than 1.75 is truncated to 1.75.

The cSEM is one of the most used values for understanding individual scores. However, many stakeholders want to have the SEM for a particular set of scores, which is:

$$SEM=\sqrt{N^{-1}\sum_{j=1}^{N}n_j\{cSEM(SS_j)\}^2} \tag{15}$$

### Total Scale Score

Earlier, this section described how the cSEM and SEM for section scale scores are obtained. The Total scale score (TSS) is a sum of the two section scale scores, and the related cSEM and SEM are computed for those composites. The cSEM for the Total scale score of the $j^{th}$ examinee is:

$$cSEM(TSS_j)=\sqrt{[cSEM(RWSS_j)]^2+[cSEM(MSS_j)]^2} \tag{16}$$

The SEM for the Total scale score can be computed using Equation 15 and replacing $cSEM(SS_j)$ with $cSEM(TSS_j)$.

## 6.3.3 Reliability

This section describes the methods to estimate reliability for the two-section scale scores, the Total score, and the three essay dimension scores.

### Scale Scores

Reliability estimates are derived for reported scale scores using the information computed in Equation 16, along with the variance of the observed scale scores. Each section and the Total score have their reliability estimates defined as:

$$\hat{\rho}_{SS}=1-\frac{SEM_{SS}^2}{\hat{\sigma}_{SS}^2} \tag{17}$$

where SS represents one of the two-section scale scores ($SS$) or the Total score and $\hat{\sigma}_{SS}^2$ is the observed variance in the scale scores of interest.

## Essay Scores

Estimates of reliability for the three essay dimension scores focus on the extent to which two independent raters agree on the score assigned. The first estimate is a simple percentage agreement, expressed as the number of agreements divided by total observations:

$$p = \sum p_{ij} \quad for\ all\ i = j \tag{18}$$

A shortcoming of simple percentage agreement is a tendency to overestimate the level of agreement (Hallgren, 2012) because it does not consider agreements due to chance.

An estimate of a single rater reliability coefficient $\hat{\rho}_{RR'}$ for a given essay dimension, the score can be estimated as the Pearson correlation between the first and section rater scores. An estimate of the single-rater SEM is:

$$SEM_R = \sqrt{\sigma_R^2 (1 - \hat{\rho}_{RR'})} \tag{19}$$

where $\sigma_R^2$ is the single-rater variance for a dimension score using the variance of ratings by both rates and is given by:

$$\sigma_R^2 = \frac{\sigma_{R1}^2 + \sigma_{R2}^2}{2} \tag{20}$$

Two additional indices of rater agreement consider agreement due to chance. The first is Cohen's kappa (Cohen, 1960), or simple kappa, which simply considers chance agreement, which is defined as the marginal frequencies of each rater's score. The measure can range from -1 to 1, where 1 is complete agreement, 0 is random agreement, and -1 is complete disagreement, and is given by:

$$\hat{\kappa} = \frac{p_O - p_E}{1 - p_E} \tag{21}$$

where $p_O$ is the observed probability of agreement and computed as in Equation 18, pE is the expected probability of agreement and is computed as $p_E = \sum p_{i.} p_{.j}$ for all $i = j$.

The last measure is the weighted kappa statistics, which considers chance agreement and penalizes disagreements. The weights used as the penalty are based on the magnitude of disagreement. The measure has the same range and interpretation as the simple kappa. The index is given by:

$$\hat{\kappa}_w = \frac{p_{O(w)} - p_{E(w)}}{1 - p_{E(w)}} \tag{22}$$

where $p_O(w)$ is the probability of agreement and computed as $p_{O(w)} = \sum \backslash no \lim its_i \sum \backslash no \lim its_j w_{ij} p_{ij}$, and pE(w) is the expected probability of agreement and is computed as $p_{E(w)} = \sum \backslash no \lim its_i \sum \backslash no \lim its_j w_{ij} p_{i.} p_{.j}$. The weights, $w_{ij}$, are constructed so that $w_{ij} = 1$ for all $i = j$, $0 \le w_{ij} < 1$ for all $i = j$, and $w_{ii} = w_{jj}$.

## 6.3.4 Analysis of Score Structure

The methods of analyzing the factor structure for each section of the SAT are described below. Given the multi-stage testing and adaptive framework for the new updated exam, two approaches are used to establish the defensibility of specifying a single factor for Math and a single factor for Reading and Writing. The two approaches are confirmatory factor analysis and generalizability theory.

### Confirmatory Factor Analysis

Confirmatory factor analysis (CFA) is one way to provide evidence of the unidimensionality of a set of questions or test items that are designed to represent a particular construct. Having an a priori specification of how items are proposed to relate to each other differentiates the method from a pure exploratory factor analysis (EFA). The goal is to evaluate whether it is reasonable that test questions that have been assembled to measure one domain constrain the measurement model to a single factor. Therefore, the dimensionality of the Reading and Writing items and Math items is examined separately. Analyses are repeated for each of the available routes students may take based on the assembled modules. Analyzing over routes is necessary due to the MST framework.

For Math, there are 20 items that contribute to the routing module, and 20 items that contribute to the stage two module. The 40 items are loaded onto a single factor, with variance of the latent factor specified to be 1. For Reading and Writing, there are 25 items that contribute to the routing module, and 25 items that contribute to the stage two module. The 50 items are loaded onto a single factor, with variance of the latent factor specified to be 1.

Measures of fit for the CFA models are: comparative fit index (CFI), Tucker-Lewis index (TLI) (also Bentler and Bonnet non-normed fit index, or NNFI), standardized root mean square residual (SRMR), and the root mean squared error of approximation (RMSEA) and accompanying 90% confidence interval. Broadly categorized, the CFI and TLI can be described as incremental fit indices, and the SRMR and RMSEA can be described as error of approximation indices (Marsh et al., 2005). Given that the investigation of score structure is a confirmatory approach, primary attention will be given to the error of approximation indices (RMSEA and SRMR). Measures of incremental fit (CFI and TLI) will be given for completeness.

In terms of evaluating the RMSEA and SRMR, indices less than 0.05 indicate good model fit. Values that are between .05 and .08 indicate acceptable model fit. In terms of evaluating the CFI and TLI, values larger than .95 indicate good model fit. Values that are between 0.90 and 0.95 indicate acceptable model fit. Current results indicate that virtually all models satisfy the RMSEA and SRMR criteria. The results for CFI and TLI are more varied, with some models falling below the desired .90 range. It should be noted that in every case, a low incremental fit value is accompanied by an on-target approximate fit value.

## 6.4 Item Analysis, Calibrations, and Pre-Equating

Once parent model templates and associated generated items have passed content, accessibility, and fairness reviews, the items are eligible to be pretested. The original pretesting for the digital versions of the assessments was completed in standalone pretesting events. With the launch of the digital assessments, starting in October 2023, items to be pretested are embedded within each section of the assessment. Pretesting occurs only within the adaptive MST. Linear versions of the test are intended for low-volume use, and to keep the length and time of the test to a minimum, it was decided not to include pretest items in the linear versions.

Within each test section (i.e., Reading and Writing and Math) of the MST, a student will see four embedded pretest items. Two will appear in stage one of each section and two will appear in stage two of each section. Pretest items are randomly assigned to examinees. Pretest items may appear in any location, subject to some content considerations, except as the first or last two items within a stage. Random assignment of pretest items to position reduces context effects and allows items to be used anywhere within future operational panels. Pretest items assigned to the second stage will be the same across both levels of module difficulty for a given student, which helps ensure that pretest items appear in all routes and will be seen across the entire ability distribution. Because the MST version of the test is the primary mode students will take the assessments in, pretested items will be seen by a broad representation of students, including those with accommodations, those testing in school day settings, and multilingual learners.

The three assessments of the SAT Suite share a common pre-equated item pool. Items are pretested across the SAT and PSAT/NMSQT and PSAT 10 assessments. At a minimum, every pretested item will be administered to 1,000 examinees and the pretested items from the same parent model will be combined to be seen by a minimum of 5,000 examinees. The precise number of generated items varies from parent model to parent model and these values are absolute minimums needed to conduct pretest item analysis. For test security purposes, pretest items are only administered to domestic U.S. test takers. For international test takers, anchor items are placed in the pretest slots, so all students taking the MST see the same number of total items regardless of whether they are administered pretest items or not. Anchor items in this context refer to items that “anchor” the other items to the IRT metric of the item pool, allowing additional statistical analyses to be conducted but are not new pretest items. However, these anchor items also do not count toward the examinee’s score.

The analysis of pretest items begins once certain conditions are met, including minimum sample sizes overall as well as key subgroups, and when most irregularity and test security issues are resolved. If students’ scores are on hold due to test security or other irregularities, those students are removed from the dataset used to analyze pretest items.

The next step of the analysis of pretest items is to compute standard descriptive statistics for each pretested item. These statistics include item difficulty (i.e., percent answering the item correctly), item discrimination (i.e., item–total score correlations where total score is the scale score reported to the student on the operational items), and distractor analysis. Pretest items are flagged for review when item difficulty exceeds 0.90 or is lower than 0.20, discrimination is less than 0.20, or a distractor has a correlation with the total score greater than 0.05. Any flagged item is reviewed further by psychometric and test development staff. If any pretest item is identified as problematic, then the entire parent model and all related items are flagged as “do not use.”

The next step of pretest analysis is to calibrate the pretest items to evaluate the performance of those items relative to each other item arising from the parent model. Item calibration is carried out using the flexMIRT software and the IRT models described in Section 6.2.1. To aid in model convergence, several prior distributions are assumed. A normal distribution prior is specified for the estimation of ability. College Board also assumed hyper-priors for the distributions of the slope (i.e., discrimination parameter), the intercept (which is related to the difficulty parameter), and lower-asymptote parameter—all specified for the initial calibration run, but may be updated if model convergence is an issue. One exception is for the SPR items in Math, where the GRM model is applied and there is no need to set prior for the lower-asymptote parameter (see Section 6.2.1 for more information). If convergence problems arise, adjustments may be made to the hyper-priors for one or more pretest items. If convergence remains problematic, the pretest item is removed from the calibration data by coding the item as not presented across all students. Operational items administered to the student are used in the calibration process by fixing the item parameters to the values used for scoring. This process is known as fixed-item calibration. Before the fixed-item calibration, the operational item performance is evaluated to ensure these items can properly function as anchors using a procedure called anchor item evaluation, which is similar to the evaluation of linking items in traditional equating. This process places all the new item parameters onto the common IRT metric which all the parent models and associated items in the pool are on.

Once model convergence is reached, psychometric staff review item fit statistics as well as item fit plots. Other model fit statistics are reviewed to ensure the convergence criteria meets minimum thresholds, which include first- and second-order convergence as reported by the flexMIRT software. Once calibration is completed, an IRT *p*-value is computed for each pretest item, which is a sample weighted estimate of item difficulty similar to a standard item difficulty. Within a parent model, these IRT *p*-values for the sample of pretested items are compared and all must be within a range of 0.10 of each other. For any parent model that has a single pretested child item outside this range, the parent model and all related child items are flagged as “do not use,” and the items are returned to test development for further review, revisions, and re-pretesting. In cases where a parent model had been earlier flagged due to convergence issues or previous review, those parent models and related child items are also marked as “do not use” and returned to test development for further review and revision.

Unless a specific pretest item has been flagged as “do not use,” all items are retained to the next step of the item calibration process. Additional criteria are also examined, including item characteristics curves (ICC) and item information functions (IIF) to ensure the related pretest items of a parent model have similar overall ICCs and IIFs. The item response data is reorganized to aggregate the responses to all parent model–related pretest items into a single item, as evidence has indicated these items arise from the same generating parent model. For pretest items that did not meet this criterion, they are treated as unique items at this stage of the calibration process. The above-described calibration process is repeated again until model convergence is obtained. Similar to the previous steps, there may be a need to remove pretest items or parent models if the model does converge. Like the above-described process, item fit and model fit are again reviewed.

The penultimate step for pretesting before approval for operational use includes evaluating all parent models for item discrimination and differential item functioning (DIF). As described above, the item discrimination is computed for parent models and any parent model that has a negative item-total correlation is flagged as "do not use." For the remaining parent models, College Board computes DIF following the Mantel-Hanzel process and applies the classification system popularized by ETS®. Under this process, parent models can be flagged as "A" which indicates no significant DIF, "B" which indicates moderate DIF, and "C" which indicates strong DIF. Any parent model showing C-DIF for any subgroup is flagged as "do not use." Parent models not yet rejected at this point are further flagged for evaluation if certain criteria are also not met. One example, if the parent model's discrimination parameter is low, the parent model is rejected or flagged for further review by psychometric and content staff.

For parent models and related child items that do not pass pretesting analyses, they are flagged as "do not use" and sent back to the test development staff to determine whether revisions to the model could be made that would improve the performance of the items. If so, the parent model is revised, new child items are generated following the full development process as described in Chapter 3, Test Development and Assembly, and then the model is re-pretested. If it is determined by the content experts that the parent model cannot be improved, the parent model is archived. Parent models are revised and re-pretested based on an assessment of (1) whether a specific flaw or flaws in the model can be identified and (2) whether that flaw or those flaws can be productively addressed. If either prong of this test fails, the model is archived.

The final step within psychometrics is a review by senior staff to ensure all steps were followed accurately, review results, and sign off on the calibration. Once signoff has occurred, only the parent models and their related items that have not been flagged as "do not use" are approved for use operationally by psychometrics. If, for any reason, signoff does not occur, the process is revisited, and results updated to address the issue until signoff occurs.

Upon completion of the calibration process, all approved parent models and related items have been placed on the same common IRT metric. This process is also known as common-item equating to a calibrated pool (Kolen & Brennan, 2004). Having all the item pools share a common IRT metric allows for item pre-equating to be used, in which the MST panels and linear forms are constructed to strict statistical specifications.

Each SAT Essay passage is reviewed by College Board staff and by external fairness reviewers. For more information on the field testing and analysis of Essay prompts, please see Chapter 3, Section 3.1.2, The Digital SAT Essay.

## 6.5 Panel Assembly and Ongoing Psychometric Quality Management

Operational analysis for the digital SAT Suite can be divided into two aspects: 1) test assembly and delivery, and 2) ongoing quality management.

### 6.5.1 Panel Assembly

To assemble a panel or a linear version[9] of any of the assessments, College Board uses automated test assembly (ATA). This process uses linear programming methods to select a set of parent models that meets an extensive list of constraints. These constraints can be broken down into content-related constraints and statistical constraints. The constraints may be classified as required or preferred. A required constraint must always be precisely met for a panel to be assembled and used operationally. Preferred constraints typically are expressed to allow some flexibility. Preferred constraints are deemed not to impact the comparability of the assembled panels. Content constraints reflect the test blueprint, whereas statistical constraints reflect the score comparability of any assembled panel. To ensure maximally comparable scores across all assembled panels, College Board implements two constraints (see Chapter 2, Section 2.3) on the statistical characteristics of every assembled panel, specifically test information functions (TIF) and test characteristic curves (TCC). Across all constraints, previous and expected item exposure are considered when selecting parent models for a panel. In considering item exposure in assembling each panel, College Board explicitly controls for the exposure of parent models and child items through the number of panels assembled, item reuse rules, and targeted rates of exposure. The overall targeted exposure rate for a child item is a maximum of 5%.

TIFs are utilized within an IRT framework and are related to reliability and standard error of measurement (see Section 6.3). An essential requirement to score comparability is that the various assessment versions have similar reliability estimates. In the context of IRT, this is achieved by requiring that the TIFs meet a specific target. Two constraints are placed on the TIFs. First, the ATA must minimize the difference between the target TIF and the TIF estimated from the selected items. This minimization is never perfectly achieved, so the second constraint places a bound around the estimated TIF, which requires the estimated TIF to be within the specified range. The range is based on research that shows scores arising from panels meeting these constraints result in scores that have similar reliability.

The constraint on TCCs is equivalent to traditional concepts of equating. The digital SAT Suite of Assessments is pre-equated, meaning that equating after an administration is no longer performed, but instead all equating is done through pretesting calibrations and linking to a calibrated item pool and ATA assembly. Like the constraints placed on TIFs, the TCCs are minimized toward a target but have a fixed range that all panel TCCs must fall within. These ranges were based on research, so all panels meeting this requirement would produce similar score

9 For ease of reading, the remainder of the section will refer to "panel," but the concepts apply equally to linear versions of the assessment.

distributions. It would be a matter of indifference to students regarding which panel they are assigned. If a panel does not meet either of these requirements, it is rejected and not further evaluated.

Once panels are assembled, a simulation study is conducted using a College Board proprietary system. The simulation study generates simulated responses to each item on a given panel thousands of times across the entire range of abilities. The ability distribution is modeled to reflect the distribution of examinees that typically take the assessment. The process is repeated multiple times for each panel to ensure the simulation results are not statistically biased. The results of each panel simulation include estimations of conditional standard errors of measurement, overall standard error of measurement, reliability, routing accuracy, and projected score distributions. To approve a panel for operational use, a panel must have an estimated reliability of 0.90 or greater for each section score, appropriate routing accuracy given the projected student population, and the score distributions across all panels to be used must be similar.

## 6.5.2 Ongoing Psychometric Quality Management

The second aspect of operational analysis relates to ongoing psychometric quality management. Depending on the assessment, a repertoire of test security analytics (see Chapter 4, Section 4.3) will be performed either on a rolling basis (window-based events) or immediately afterward (weekend-based events). In addition, other monitoring—online and offline—of test irregularities may lead to potential holds of score release, further investigation, and a potential hold on score reports. Beyond scoring quality control (see Chapter 5, Test Scoring and Reporting) and test security, ongoing monitoring of scores will be performed at regular intervals to identify abnormal patterns in score distributions to prevent potential issues in reported scores. As College Board worked toward launching the digital SAT Suite, a repertoire of operational analyses was developed, which were first applied to the March 2023 administration and are being modified and improved as College Board accumulates more administration data. College Board's approach to quality management will include continuous improvement to the production process so that the analyses conducted may be added to, removed, or refined. More traditional item and test analyses, some of which are characteristic of the multistage adaptive design, are in place once adequate score data for targeted items and panels are collected from operational administrations, including but not limited to the following categories:

- Classical and IRT analysis on operational parent models and panels
- Item difficulty and discrimination
- IRT model fit evaluation
- IRT parameter and scale drift
- Reliability and SEM/cSEM
- Evaluation of parent model performance and child item consistency
- Score consistency across administrations
- Evaluation of adaptive testing in reference to test design and assembly
- Evaluation of empirical score distributions
- Routing decisions and impacts
- Monitoring of internal structure of assessments

The above analyses will be performed in the context of historical trend analysis to identify items displaying scale drift over time, conditional on student ability. Changes in item performance, as well as parent model drift, are monitored. For example, suppose item performance or parent model is identified as aberrant over time or lacking fit to the IRT model. In that case, it will be flagged for further review and the item will be evaluated and considered for recalibration or possible removal from the item pool or reviewed by test development staff.

On a more general level, monitoring item and panel performances will inform the evaluation of the stability of score scales used in reporting, the appropriateness of the psychometric model, and the specifics of item calibration and ability estimation. In light of such evaluations, the overall trend in item performance will also be assessed by College Board test development groups to adjust their strategy and prioritization towards item generation, pretesting, and item banking to sustain the quality and consistency of test scores over time. This information will also be used to refine test assembly and other processes. Psychometric operational analysis is not used to inform administration material revisions.

Chapter 7

# Fairness

## 7.0 Introduction

This chapter discusses issues of fairness and College Board's commitment to ensuring fairness for all facets of the digital SAT Suite of Assessments. Section 7.1, The Concept of Fairness in Testing, provides a brief overview of the concept of fairness in testing, spotlighting the principles that remain front of mind during the development and administration of the SAT Suite. Section 7.2, Fairness Reviews of Test Materials, looks at the rigorous multifaceted review process employed by College Board to ensure that the tests of the digital SAT Suite are fair and accessible to all students. Section 7.3, Accessibility, examines the important issue of accessibility and how it is addressed by the digital test, including the design of the test and accommodations offered during testing. Section 7.4, Test Administration, covers how concepts of fairness are applied to test administration and how the transition to digital has led to many advances in this area. Section 7.5, Test Security, analyzes relevant risks and discusses how adequate security in the creation, handling, and delivery of test materials contributes to both the reality and perception of test fairness. Lastly, Section 7.6, Additional Fairness Mechanisms, provides a look at some additional mechanisms employed by College Board that contribute to ensuring that the digital SAT Suite tests are fair and accessible to all students.

## 7.1 The Concept of Fairness in Testing

College Board is strongly committed to the indivisibility of the concepts of test validity (i.e., that a test is measuring what it is intended to measure) and test fairness (i.e., that a test affords an equal opportunity to all test takers to perform up to their level of achievement without hindrance). To put the matter simply, a test must be fair to be valid. Test fairness considerations permeate the design, development, and administration of the digital SAT Suite.

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 2014 edition of the *Standards for Educational and Psychological Testing*, call for test publishers to "minimize barriers to valid score interpretations for the widest possible range of individuals and relevant subgroups" (AERA et al., 2014, p. 63). An assessment should be built in such a way that the constructs (concepts) 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* (AERA et al., 2014), are addressed as part of the development and administration of the digital SAT Suite:

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" (p. 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" (p. 51) when accommodations and supports in testing do not compromise the construct (e.g., reading comprehension) being measured.
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" (p. 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., math 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" (p. 52). Accommodations and supports may take such forms as providing students who have visual impairments with 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 items 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" (pp. 53–54).

College Board embraces the fairness expectations articulated by the AERA, APA, and NCME *Standards* and the overarching goal of ensuring the maximal inclusiveness, representativeness, and accessibility of its digital 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 digital 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 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' design
- Make test content as fully and as widely accessible to as many test takers as possible through 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
- 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

## 7.2 Fairness Reviews of Test Materials

College Board employs a rigorous, multifaceted review process to ensure, among other things, that the tests of the digital SAT Suite are fair and accessible to all students. This includes both internal and external reviews by trained experts prior to the materials' operational use as well as statistical evaluation of question fairness, chiefly via differential item functioning (DIF) analysis. Digital SAT Essay prompts undergo a broadly analogous internal review and external fairness review.

### 7.2.1 The Reading and Writing and Math Sections

#### Internal Review

Every parent model and child item developed for the digital SAT Suite Reading and Writing and Math sections as well as every prompt developed for the digital SAT Essay undergoes a thorough internal review that has as one of its goals ensuring that examined test materials are free from fairness-related issues, including but not limited to bias, stereotyping, and inaccurate or inauthentic portrayal of population subgroups, prior to being administered to students operationally.

The many stages of internal review involve the careful evaluation of test materials by trained experts for conformity to documented content, fairness, and editorial standards established by College Board.

With regard to fairness, College Board has developed and periodically updates its test review materials. The primary source of fairness guidelines resides in the suite's test review guides, which are used by both College Board staff and the independent consultants the organization hires to evaluate its test materials. The fairness portion of the test review guides presents a thorough description of the concept of test fairness, offers a range of considerations specific to the digital SAT Suite, and includes a checklist built from the considerations to better regulate and standardize reviews. All College Board staff involved in the creation and/or review of digital SAT Suite test materials have been trained on these guidelines, and College Board regularly engages with experts in education and test fairness to comment on and suggest updates and revisions to the guidelines.

## External Review

In addition to employing trained staff members to evaluate test materials for potential fairness concerns, College Board makes use of independent subject matter and fairness experts to assess the quality of digital SAT Suite test materials. In general terms, these experts are typically active (or recently retired) educators at the secondary and postsecondary levels; collectively possess both subject matter expertise and familiarity with the testing population in general and with specific population subgroups; and are highly diverse in terms of their own backgrounds and work experiences.

College Board engages independent experts in the fairness review of digital SAT Suite Reading and Writing and Math test materials in two main ways:

1. **Semiannual panel/form reviews.** Twice each year, College Board convenes groups of subject matter and fairness experts to review and assess sets of representative digital SAT Suite test materials. Such review committees, which may meet virtually or in person, receive in advance one or more sample digital SAT Suite panels and/or forms; are asked to submit ahead of the meeting their item-, module-, and panel-/form-level comments; and then discuss and resolve their comments with College Board test development staff and their peers. Reviewers hired for this review include experts on students who are part of various population subgroups, including Black/African American, Asian American, Hispanic/Latino, and Native American/Alaska Native students; experts on students who identify as male, as female, or as nonbinary; experts on students who are English learners; and experts on students with disabilities. The main goal of such reviews is to verify, at a high level, the alignment between the test materials and college and career readiness requirements for the purposes of broadly guiding future test development. In terms of fairness, this may include identifying item approaches or topics that committee members believe should be retained, modified, emphasized or deemphasized, or eliminated altogether.
2. **Item-level fairness review.** In addition to the semiannual panel reviews, which focus primarily on macro-level issues of digital SAT Suite test design and development, College Board employs committees of trained, independent experts to assess individual test items for fairness. All Reading and Writing items previously unreviewed externally are subject to this process, as are all Math items set in context. Reviewers, who work independently and whose qualifications are analogous to those described above, are provided with sufficient materials to assess the fairness of each Reading and Writing and Math parent model and

child item against criteria set forth in the test review guides. Parent models and child items deemed sufficiently problematic by reviewers are precluded from operational use, though faulty parent models may, at the discretion of test developers, be reedited and a sample of their children re-pretested. This review process is undertaken with sufficient frequency to ensure that all applicable items available for operational use have been so examined.

*Statistical Review*

In addition to the qualitative reviews described above, College Board performs a range of statistical analyses on test item performance to ensure that materials are performing as expected. Such analyses, particularly in the realm of differential item functioning (DIF), have important roles to play in assessing the fairness of test materials.

College Board measurement experts perform DIF analyses on pretested items. This statistical method, in brief, involves comparing samples of test taker performance to investigate whether certain defined population subgroups (e.g., males and females) seem to perform differently on given test items. Assuming the premise that two student samples of equivalent achievement should have an equal probability of answering a given item correctly, DIF analysis seeks to uncover cases in which the performance of one sample (e.g., a group of students identifying as female) differs from that of another sample (e.g., a group of students identifying as male) to an extent unlikely to have occurred purely by chance. When analysis of pretested child items identifies parent models exhibiting high levels of DIF, those parent models are discarded or, optionally, revised and a subset of their child items re-pretested. For further details, see Chapter 6, Psychometrics.

## 7.2.2 The Digital SAT Essay

Passages from previously published sources that are considered candidates for use as part of digital SAT Essay prompts are carefully vetted by trained College Board staff for suitability on a number of dimensions, including fairness. Internal reviewers evaluate each passage candidate against the same fairness criteria used for the other sections of the digital SAT Suite tests. Each prospective passage is also reviewed externally for fairness by diverse groups of independent experts knowledgeable about the student population of interest (and relevant subgroups within that population) and familiar with College Board's quality assurance criteria. Either internal or external reviews may determine that a given passage is suitable for use as-is, suitable for use with minor edits (e.g., adding a clarifying definition or bit of context to the passage's advance organizer; eliding a distracting minor detail), or not suitable for use. In addition, previously developed prompts are reevaluated internally prior to operational use to ensure that a previously acceptable topic and treatment have not become unusable in the interim due to unforeseen circumstances.

To ensure fair and comparable digital SAT Essay prompts, College Board completes psychometric analyses on the field test results of all prompts, including interrater reliability stats (including correlations) between scorer 1 and scorer 2 (and scorer 3 if applicable) for all dimensions (Reading, Analysis, Writing). The mean and the standard deviation for each prompt are compared, as are the frequency distributions by score point across each dimension for each prompt. Any prompt that is an outlier—that is, has a mean score on any dimension that is unusually high or low or that has a distribution of scores on any dimension that is significantly different from the distribution of scores on the other dimensions—is not used operationally. Demographic analyses are also conducted, including analyzing female/male mean differences by dimension scores by prompt and mean differences by race/ethnicity for dimension scores by prompt. Any prompt that demonstrates extreme differential performance, either between males and females or between White and other races/ethnicities, is not used operationally. For more information, see Chapter 3, Section 3.2.1, The Digital SAT Essay.

## 7.3 Accessibility

As indicated above, accessibility is a critical aspect of test fairness. The following subsections detail how the digital SAT Suite assessments advance the goal of maximal accessibility for all students through the application of universal design principles, the provision of universal tools, and the availability of accommodations and supports for those students who require them.

### 7.3.1 Universal Design

In designing and developing the digital SAT Suite, College Board adhered closely to the tenets of universal design generally and to the ways in which researchers and practitioners have recommended that these tenets be applied to the design, development, and administration of large-scale standardized tests. The concept of universal design (UD), which originated in the field of architecture, is intended to promote “the design of products and environments [so that they are] usable by all people, to the greatest extent possible, without the need for adaptation or specialized design” (Center for Universal Design, 1997). Universal design for learning (Rose et al., 2002) applies UD principles to education. It promotes providing multiple means of engagement, representation, action, and expression to students in order to reduce, if not eliminate, barriers to equitable educational access. Universal design for assessment, or UDA, (Thompson et al., 2002) extends UD principles specifically to large-scale assessment.

Table 7.1 displays the seven core principles of UDA, defines each briefly using language from Thompson, Johnstone, and Thurlow (2002), and indicates how the digital SAT Suite implements each principle, a topic that is expanded on in the subsequent discussion.

**TABLE 7.1** DIGITAL SAT SUITE IMPLEMENTATION OF UDA PRINCIPLES

| UDA Principle | Key Requirement(s) | Digital SAT Suite Implementation |
|---|---|---|
| Inclusive assessment population | "Opportunities for the participation of all students, no matter … their cognitive abilities, cultural backgrounds, or linguistic backgrounds" (p. 6) | Universal tools available to all students<br>Wide range of accommodations and supports available<br>Pretesting and studies that include English learners (ELs) and students with disabilities (SWD)<br>Quantitative data-gathering with special-needs populations |
| Precisely defined test constructs | "Remov[ing] all non-construct oriented cognitive, sensory, emotional, and physical barriers" to assessment (p. 8) | Carefully defined test specifications with construct definitions<br>Evaluation of potential accommodations and supports to confirm that they do not affect the constructs intended to be measured |
| Accessible, nonbiased test items | "Incorporating accessibility as a primary dimension of test specifications"; verifying lack of bias in items (p. 9) | Strong commitment to accessibility-as-validity from day one<br>Staff training in content and fairness criteria and appropriate language selection<br>External, independent reviews including race/ethnicity, gender, EL, and SWD representatives<br>DIF analyses |
| Amenable to accommodations | Ready provision of accommodations for SWD and supports for ELs that do not weaken the intended constructs | Carefully defined test construct descriptions<br>Availability of a wide range of accommodations and supports for those students who require them |
| Simple, clear, intuitive instructions and procedures | "Instructions … in simple, clear, consistent, and understandable language"; practice materials; standardized administration (p. 13) | Focus group evaluations of Math student-produced response (SPR) directions by test takers, including test takers using accommodations<br>Official Digital SAT Prep and other test familiarization opportunities, most of which are offered at no cost<br>Test Day Toolkit to support administration ease and consistency |
| Maximum readability and comprehensibility | Managing text complexity and "[using] plain language when vocabulary level is not part of the construct being tested" (p. 15) | Mechanisms for evaluating Reading and Writing text complexity and determining which vocabulary words/phrases are tested<br>Staff training in appropriate language selection |
| Maximum legibility | Legible texts; legible graphics; legible response formats | Careful selection of font and point size, with the ability to adjust via magnification<br>Well-articulated content and editorial parameters for graphics<br>Alt text for graphics as well as other accommodations<br>Straightforward response entry (with alternatives for students with disabilities needing accommodations) |

*Note.* This table was adapted from Thompson et al., 2002.

**Inclusive assessment population.** UDA-compliant tests must offer "opportunities for the participation of all students, no matter what their cognitive abilities, cultural backgrounds, or linguistic backgrounds" and "need to measure the performance of students with a wide range of abilities and skill repertoires, ensuring that students with diverse learning needs receive opportunities to demonstrate competence on the same content" (Thompson et al., 2002, p. 6). The digital SAT Suite meets these goals, in part, by making universal accessibility tools, such as the abilities to zoom in and to adjust contrast, available in Bluebook for all students to use or not use at their discretion; by offering a wide range of accommodations and supports to those students who require them; by including members from special-needs populations, such as English learners (ELs) and students with disabilities (SWD), in pretesting and other studies; and by engaging directly with special-needs populations via studies targeted at better understanding their requirements and preferences.

**Precisely defined test constructs.** Carefully articulating constructs—definitions of the concepts (e.g., reading comprehension, math achievement) a given test is intended to assess—promotes fairness and accessibility by differentiating between the skills and knowledge that are appropriate to assess (i.e., construct-relevant factors) and confounding elements that may impair an accurate assessment of those skills and knowledge (i.e., construct-irrelevant factors). Such construct-irrelevant factors for the digital SAT Suite include, but are not limited to, students' race/ethnicity, gender and sexual identities, home region and home culture, and whether students live in rural/small-town, suburban, or urban areas.

As Thompson et al. (2002) put it, "Just as universally designed architecture removes physical, sensory, and cognitive barriers to all types of people in public and private structures, universally designed assessments remove all non-construct-oriented cognitive, sensory, emotional, and physical barriers" (p. 8). The digital SAT Suite meets these goals, in part, by developing and publishing such constructs (see Chapter 2, Section 2.2.1, The Reading and Writing Section, and Section 2.2.2, The Math Section), by routinely engaging both internal and independent subject matter and fairness experts in assessing test content for construct relevance, and by evaluating potential accommodations and supports to confirm their efficacy and that their use does not compromise the constructs intended to be measured.

**Accessible, nonbiased test questions.** For tests to be UDA compliant, test makers must "[incorporate] accessibility as a primary dimension of test specifications" (Thompson et al., 2002, p. 9, citing Kopriva, 2000) and verify that test questions are free from bias (i.e., construct-irrelevant factors that may influence test performance in unintended ways).

College Board test development staff are carefully trained on exacting content and fairness requirements, to which they refer when developing and reviewing test materials. (See Chapter 3, Test Development and Assembly, for more information on the test development process.) Questions on the digital SAT Suite are also routinely subjected to external review for content soundness and fairness by independent educators at the secondary and postsecondary levels, who themselves follow the same content and fairness criteria that College Board staff make use of. In addition, College Board measurement experts perform DIF analyses on pretested questions, a process discussed in Section 7.2.1, The Reading and Writing and Math Sections.

**Amenable to accommodations.** According to Thompson et al. (2002), UDA-compliant tests must make ready provision of testing accommodations that allow students who need them to participate fully in the testing experience. This notion can also be extended to students whose first or best language is not English and who require supports to fully access the test content.

To these ends, the digital SAT Suite programs offer a wide range of accommodations and supports for students who require them. These offerings have been carefully evaluated to make sure that they aid the students making use of them and do not undermine the constructs the tests are intended to measure. For example, magnification devices are acceptable accommodations on the digital SAT Suite because visual acuity is not a construct being measured; however, because English language proficiency is a requirement throughout the tests' design, students are not permitted access to conventional dictionaries to look up the meaning of words and phrases used on the test, although, in some circumstances, English learners are able to use approved word-by-word bilingual glossaries as an accessibility support. The score reports of students who make use of accommodations and supports are not differentiated from those of students not using them because such assistance serves only to level the playing field between these two groups and to give students with disabilities as well as English learners equivalent opportunities to demonstrate their achievement through the tests.

**Simple, clear, intuitive instructions and procedures.** Thompson et al. (2002) call for test instructions and procedures that are "easy to understand, regardless of a student's experience, knowledge, language skills, or current concentration level" (p. 13). The authors also signal the value of providing students with practice opportunities prior to testing and making sure that test administration conditions are well documented so that they can be standardized and consistently replicated.

The digital SAT Suite furthers these goals in several ways. First, test instructions and procedures have been kept to a clarifying minimum and are available to students in advance of testing via official practice tests produced by College Board and other test familiarization opportunities (see also Chapter 4, Section 4.2.3, Accessibility). Procedures for responding to digital SAT Suite items are also straightforward: students must either select the best option among four for multiple-choice items (the format used for the vast majority of digital-suite items) or, for select items in Math, generate and enter their own answers in a format referred to as student-produced response, or SPR.

Math SPR directions have been crafted carefully by both content and user experience experts to emphasize clarity and minimize the likelihood of student entry error. Student entries in these fields are automatically validated by Bluebook to preclude certain kinds of errors (e.g., entry of too many digits; use of nonnumerical characters other than the negative sign and the slash for fractional answers), and Bluebook previews entries for test takers, allowing them to confirm that the answers they enter are the ones they intend to provide. College Board has also conducted focus groups with test takers, including those using accommodations, on their perceptions and level of understanding of the SPR directions and used this feedback to inform refinements. Finally, the Math SPR instructions, like all digital SAT Suite test directions, are available to students at any time during testing.

Second, as is detailed in Chapter 4, Section 4.2.3, Accessibility, College Board makes numerous opportunities available to students, most of them at no cost, to practice productively for the digital SAT Suite tests.

Third, the digital SAT Suite represents a significant enhancement of test administration ease over its paper and pencil predecessor, thereby improving standardization of test delivery. Test Day Toolkit, College Board's test administration app, vastly simplifies the task of giving the digital SAT Suite tests by guiding proctors through every step of the process. In addition, test timing—formerly a key responsibility of test proctors—has been transferred to Bluebook, thereby ensuring that all students have the exact time allotted for testing and have ready access to accurate information about how much time is left for a given test module. These innovations make it even more likely that every student taking one of the digital SAT Suite assessments does so under precisely the same conditions (with the important exception that approved accommodations and/or supports may intentionally alter testing circumstances as a means to promote test equity).

**Maximum readability and comprehensibility.** Thompson et al. (2002) advocate that test makers control the linguistic demands of test questions to ensure that texts are no more complex than they need to be to satisfy the demands of the construct being measured and that items "use plain language when vocabulary level is not part of the construct being tested" (p. 15).

The digital SAT Suite approaches these requirements in two main ways. First, because text complexity is an important aspect of the reading and writing construct being assessed, College Board staff use a mixture of both quantitative and qualitative measures of text complexity to ensure that the passages used as stimuli in test items represent appropriate levels of challenge consistent with college and career readiness requirements. On the quantitative side of the equation, College Board has developed a custom text complexity measurement tool that is more suitable than off-the-shelf products for assessing accurately the difficulty of the brief Reading and Writing passages used on the tests. Because no quantitative tool, no matter how well designed, can consider all the factors that contribute to text complexity, College Board requires staff trained on text complexity requirements to also weigh in on judgments about which of the digital suite's three text complexity bands—grades 6–8, grades 9–11, and grades 12–14—a given text is most appropriate for.

Second, digital SAT Suite test development staff have been trained on principles of linguistic modification (Abedi & Sato, 2008), a set of practices designed to ensure that the language used in test materials is as broadly accessible as possible and is only as complicated as it needs to be to assess the constructs being measured. This training and these approaches are particularly important for the Math section, for which text complexity is not a formal requirement and for which maximal transparency in and clarity of linguistic expression is therefore necessary to prevent extraneous elements of language difficulty from impeding students' performance on the construct.

**Maximum legibility.** Finally, Thompson et al. (2002) set forth the expectations that text and any graphics presented to students be clear and legible and that response formats be designed with the needs of all test takers, including those with visual impairments or issues with fine motor skills, in mind.

For the digital SAT Suite, College Board has selected a typeface—Noto Serif 15/24—that displays text clearly on a range of digital devices and screen sizes. Via a universal tool (i.e., one not requiring an approved testing accommodation), all students can use built-in magnification to increase display size. Well-articulated content and editorial standards govern the development and presentation of informational graphics, such as tables and graphs, needed as stimuli for select items; for students requiring verbal descriptions of graphics to respond to test items, alt text has been created that allows students to access the test content in ways that do not disclose the intended answer to a given graphics-based item. More generally, alternate test formats and response modes, both digital and paper based, are also available for students who require them to participate fully in the tests. For all other students, the universally designed response formats—multiple-choice and student-produced response modes—have, as discussed previously, been carefully vetted for ease of use and minimization of student entry error.

### 7.3.2 Universal Tools

Bluebook supports a number of universal tools that all students, at their discretion and preference, may use or not use to improve their test-taking experience. These tools include a built-in version of the Desmos Graphing Calculator, an annotation tool, an answer choice elimination tool, and a method of marking questions to be reviewed before time elapses. By design, some of the universal tools the digital test utilizes, such as those for zooming in and out, were previously offered only as accommodations (e.g., large print); their universal availability in the digital SAT Suite serves to increase the accessibility of the tests for all students.

### 7.3.3 Accommodations and Supports

#### Accommodations for Students with Disabilities

To ensure that a fair testing environment is available to all test takers, College Board provides students with disabilities taking the digital SAT Suite assessments with the accommodations that they need. This practice ensures that, via the approved accommodations, College Board is removing or minimizing construct-irrelevant barriers that can interfere with a test taker accurately demonstrating their true standing on a construct (AERA et al., 2014).

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 AERA, APA, and NCME *Standards* and industry 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 et al., 2014, p. 67). To this end, all accommodated test forms and testing conditions are designed to be comparable, in that even though forms or conditions might be modified based on the needs of a particular test taker, the constructs being tested and the meaning of the scores obtained remain unchanged.

Although numerous accommodations are possible, students with disabilities must submit a request for approval by College Board. The vast majority of students who are approved for and using testing accommodations at their school through a current Individualized Education Program (IEP) or 504 Plan have those same accommodations approved for taking 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 in which a student does not qualify for automatic approval through the school verification process, the request and documentation are reviewed by College Board's Services for Students with Disabilities (SSD) department. 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 impairments, learning disorders, physical and medical impairments, and motor 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 accommodation(s) 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 digital SAT Suite and AP Exams. Students do not need to request accommodations from College Board for subsequent assessments taken during this eligibility period unless their accommodations needs change. More information about the availability of accommodations and the procedures for requesting them prior to testing can be found at College Board's SSD website, **collegeboard.org/ssd**.

## Commonly Offered Accommodations

The following is a list of accommodations commonly offered as part of the digital SAT Suite. Accommodations are not limited to those listed, as College Board considers any reasonable accommodation for any documented disability as long as a student qualifies for testing accommodations.

### Timing and Scheduling

- Extended time: Time and one-half (+50%), double time (+100%), more than double time (>+100%)
- Extra/extended breaks

### Reading/Seeing Text

- Text to speech
- Magnification device (electronic or nonelectronic)
- Color contrast
- Braille with raised line drawings (Unified English Braille [UEB] Contracted and Nemeth Braille Code)
- Raised line drawings
- Braille device for written responses
- Other: Reading/seeing text

### Recording Answers

- Writer/scribe to record responses
- Braille writer
- Dictation/speech to text
- Other: Recording answers

### Modified Setting

- Small-group setting
- Preferential seating
- Wheelchair accessibility
- School-based setting
- One-to-one setting
- Other: Modified setting

### Other

- Food/drink/medication
- Permission to test blood sugar
- Sign language interpreter (for oral instructions only)
- Printed copy of verbal instructions
- Assistive technology
- Auditory amplification/FM system
- Other

## Paper-Based Testing

College Board's digital testing app Bluebook was designed expressly to meet the accessibility needs of the vast majority of digital SAT Suite test takers, including students with testing accommodations. As the preceding subsection notes, Bluebook supports a wide range of accommodations, and the provision of a set of universal tools, such as magnification adjustment, further obviates the need for some kinds of specialized accommodated test forms.

At the same time, some students, for various reasons, still need to test on paper to have unfettered access to the digital SAT Suite tests. To meet this need, College Board develops paper-based versions of the assessments for those students who require them. The overall design of the paper test forms closely mimics the content and statistical specifications of the multistage adaptive versions, the chief differences being that the paper-based test forms are not adaptive in nature and thus are somewhat longer in terms of number of items and time allotted. The scores reported for the paper-based test forms are identical to those reported for the digital adaptive forms and are college reportable.

## Supports for English Learners

To better serve students who are acquiring English, College Board offers testing supports for ELs during SAT School Day, PSAT 10, and PSAT 8/9 administrations. Supports are not currently available for weekend SAT administrations or for the PSAT/NMSQT. EL supports will be available for PSAT/NMSQT starting fall 2024. Any student who currently uses the supports for PSAT/NMSQT will receive scores for guidance purposes only.

Testing supports include the following:

- **Translated test directions.** Directions are available in English and 14 other languages for the digital SAT, PSAT 10, and PSAT 8/9.
- **Use of bilingual word-to-word dictionaries.** The bilingual word-to-word dictionaries students use on test day must be from a College Board–approved list. Schools will provide the dictionaries to students on test day, collecting them when testing is complete.
- **Time and one-half (+50%) extended testing time.** When EL students use extended time on test day, they are given time and one-half (+50%) on each test section. Students using time and one-half (+50%) for EL purposes may test in the same room(s) as other students using time and one-half (+50%) for the full test. Extended time for EL students can only be used on the test date indicated; unlike accommodations for students with disabilities, EL supports are temporary.

EL students who use one or more of the above supports during the digital SAT will receive scores they can send to colleges.

Students who meet the following criteria at the time of testing can use EL supports:

- They are enrolled in an elementary or secondary school in the United States or U.S. territories.
- They are an English learner as defined by their state or by federal policy.
- They use the same supports in class or for other assessments.

More information about the availability of supports and the procedures for requesting them prior to testing can be found at College Board's SSD website, **collegeboard.org/ssd**.

## 7.4 Test Administration

An important plank in the definition of *fairness* provided in Section 7.1, The Concept of Fairness in Testing, is that standardization in test administration is a traditional and important part of ensuring that all test takers have fair and equitable opportunity to demonstrate what they know and can do with respect to the constructs of interest. At the same time, the *Standards* (AERA et al., 2014) call for necessary flexibility in such standardization through the provision of appropriate testing accommodations and supports that do not compromise the constructs being measured.

While College Board has always striven to ensure that its SAT Suite tests are as consistently administered as possible, the transition to digital adaptive testing has allowed for important advances in this area. The two key technological innovations are the use of Bluebook, a customized test delivery application, and Test Day Toolkit, an app for test proctors and test center coordinators.

### 7.4.1 Bluebook

Bluebook has replaced the paper-based assessment model for the vast majority of digital SAT Suite test takers (the exception being for those students who require paper-based testing as an accommodation). Among its many features and relevant to the present discussion is that Bluebook, and not the test proctor, handles test timing. This ensures that every student taking one of the digital SAT Suite tests has exactly as much testing time as every other student using the application.

### 7.4.2 Test Day Toolkit

Test Day Toolkit makes every aspect of giving the digital SAT Suite assessments simple and easy, with concomitant gains in test administration standardization. For all users, Test Day Toolkit facilitates student check-in on test day, allows for the customization of seating charts, and offers a straightforward process for submitting testing irregularity reports online. Test center coordinators have additional tools, such as the ability to set up testing room assignments automatically, print rosters, and report test center closures. By handling all these (and other) tasks in an efficient, streamlined way, Test Day Toolkit enhances the standardization of test administration and thus promotes the goal of test fairness for all students.

## 7.5 Test Security

Although not formally addressed in the four-part definition of *fairness* laid out in Section 7.1, The Concept of Fairness in Testing, test security is a crucial prerequisite to achieving the other listed aims. Without adequate security in the creation, handling, and delivery of test materials, both the reality and perception of test fairness are severely threatened.

The transition to the digital SAT Suite yielded numerous benefits for test security.

- **Unique test forms.** Each test taker is administered a highly comparable but unique test form. This greatly reduces the potential value to students of both illicitly acquiring previously administered unreleased test materials and attempts to gain an unfair advantage or collude in a test room environment.

- **Greater item variety.** Because the item pools for the digital SAT Suite assessments are so large (see Chapter 3, Section 3.1, Test Content Development), the relative value to bad actors of any leaked content is substantially reduced relative to breaches involving the theft or duplication of entire paper test booklets.
- **Limited content exposure.** Bluebook displays each test item (and its accompanying stimulus) separately, thus greatly impairing the ability of bad actors to copy test materials efficiently and without being observed by test proctors.

## 7.6 Additional Fairness Mechanisms

Two additional mechanisms employed by College Board contribute to ensuring that the digital SAT Suite tests are fair and accessible to all students.

### 7.6.1 Student Postexperience Surveys and Focus Groups

Since 2021, College Board has systematically collected feedback from student test takers on various aspects of the digital-suite experience, including students' perceptions of the comprehensibility and ease/difficulty of the test items, the quality of the opportunity offered to demonstrate their skills and knowledge, their academic preparedness for and comfort with answering the test items, the tests' timing conditions, and their experiences with the digital test delivery interface. College Board has also inaugurated a series of ad hoc student focus groups to obtain more systematic feedback on the digital-suite tests. These focus groups to date have included both broad cross sections of the SAT Suite test-taking population as well as groups composed of members of specific test-taking subpopulations, including English learners and international students. Input from these surveys and focus groups has been correlated with student performance data to help evaluate the test design and identify potential refinements to the tests. College Board will continue to conduct surveys and meet with focus groups in order to collect feedback directly from the tests' most important users: students themselves.

### 7.6.2 Test Materials Challenge Procedure

To further promote test transparency, College Board continues to make available to test takers and proctors methods for reporting concerns about test materials to the organization. Students may report any concerns they have about the accuracy, correctness, or appropriateness of test items they are administered to their proctors or directly to College Board via established channels. In the rare event that a problem is uncovered with a test item during its evaluation process, College Board will take additional steps, up to and including (1) determining whether an item is flawed and should not be scored and (2) ensuring that any flawed item does not appear in its present form in any future operational materials.

Chapter 8

# Validity

## 8.0 Introduction

This chapter covers validity as it pertains to the digital SAT Suite. Ensuring that test scores and interpretations are supported by strong validity evidence in accordance with their intended purposes and uses is a key component of creating and maintaining high-quality assessments. As such, a commitment to matters of validity is of paramount importance as College Board develops, administers, and scores the digital SAT Suite tests. In many ways, each preceding chapter in this manual covers validity, as all the processes and procedures described heretofore have as their main goal the generation of digital SAT Suite tests and scores that are valid for their intended purposes and uses. A deeper examination of the issues surrounding validity is placed here at the end of this manual, as first acquiring an understanding of the many processes and procedures involved in creating, administering, and scoring the digital SAT Suite tests, as well as how and why their scores are intended to be interpreted and used, allows the reader to more fully comprehend the wide-ranging validity evidence herein presented.

Our examination of validity as it relates to the SAT Suite begins in the broadest terms possible, as Section 8.1, Introduction to Validity as a Concept, provides a brief overview of validity and the goals of test score validation. Section 8.2, Content-Oriented Validity Evidence, presents the evidentiary foundations undergirding the content of the digital SAT Suite tests. Section 8.3, Response Process–Oriented Validity Evidence, describes studies that confirm that items on the digital SAT Suite are capable of eliciting from students the sorts of higher-order, cognitively complex thinking required for college and career readiness, as well as studies conducted by College Board to evaluate and improve on the more complex of the two response formats used for digital SAT Suite test items and the Bluebook application itself. Section 8.4, Relationship Between the SAT and Other Educational Measures, documents the strong, positive relationships between digital SAT Suite test scores and other educational measures of academic achievement, while Section 8.5, Relationship Between the SAT and College Outcomes, analyzes how digital SAT Suite scores relate to the first-year college outcomes they are intended to predict. Finally, Section 8.6, Measuring and Monitoring College and Career Readiness with the Digital SAT Suite, looks at how College Board's empirically derived benchmark scores provide a way to monitor student progress toward college and career readiness from year to year.

# 8.1 Introduction to Validity as a Concept

Gathering validity evidence is one of the most important steps in creating and understanding test scores and their uses. As per the AERA, APA, and NCME *Standards*, *validity* is defined as “the degree to which evidence and theory support the interpretations of test scores for proposed uses of tests” (AERA et al., 2014, p. 11). This means that if a test score has multiple uses and interpretations (e.g., admission to an institution and placement into coursework), each distinct use and interpretation must be validated. Note that validity is not a property of the assessment itself but rather refers to the interpretation of test scores for a specific use.

The goal in test score validation is to develop a logical rationale for the proposed uses of test scores and then to find evidence to support (or refute) those uses (Kane, 1992, 2006, 2013; Sireci, 2013). For the SAT Suite, this validity argument would include the evidence from new studies and some previously reported research on the utility of scores for the purposes of evaluating and monitoring students’ college and career readiness and making college admission decisions.

The validity evidence-gathering process can be thought of as an iterative, never-ending process to help improve the assessment for each use (Kane, 2013). The more evidence collected supporting a score’s particular use, the stronger the argument for that particular interpretation of the test score. Validity evidence for the SAT Suite is gathered in multiple ways. Evidence includes (but isn’t limited to) consideration of the content of the assessment (e.g., subject areas covered and the format of the items), the internal structure of the assessment (e.g., psychometric properties), and how test scores relate to other variables (e.g., predictive relationships).

This chapter outlines and discusses the validity evidence gathered to support the proposed uses and interpretations of the SAT Suite. As the AERA, APA, and NCME *Standards* (2014) note, “Fairness is a fundamental validity issue and requires attention throughout all stages of test development and use” (p. 49). Test fairness and related considerations and analyses are foundational to the assessment design and, as such, are an integral aspect of all SAT Suite validity research.

## 8.2 Content-Oriented Validity Evidence

### 8.2.1 Curriculum Survey

College Board content and measurement staff made extensive use of curriculum survey data to inform decisions about which skills and knowledge should be tested on the digital SAT Suite. College Board's most recent curriculum survey data (College Board, 2019) were collected from (1) a nationally representative sample (n=1,645) of postsecondary faculty at two- and four-year institutions who teach courses in English, math, social science, and science, and (2) a nationally representative sample (n=2,686) of middle school/junior high school math teachers and high school English language arts and math teachers.

College Board analyzed the collected data to answer two main questions:

1. To what extent are the English language arts/literacy and math skills and knowledge measured on the SAT Suite deemed important for incoming students to have attained in order to be ready for and successful in entry-level, credit-bearing two- and four-year postsecondary English, math, social science, and science courses?
2. To what extent are the skills and knowledge measured on the SAT Suite being taught in middle school/junior high school math, high school math, and high school English language arts classrooms?

For the present purpose, the answer to question 1 is more central because it relates directly to the core purpose of the SAT Suite: measuring students' attainment of essential college and career readiness prerequisites.

Although this curriculum survey study was framed with the paper-based SAT Suite and its specifications in mind, its data nonetheless support the conclusion that the digital SAT Suite, which, by design, measures a highly similar range of skills and knowledge, also addresses critical college and career prerequisites in English language arts/literacy and math.

Table 8.1 and Table 8.2 summarize the key findings relevant to the digital suite in English language arts/literacy and math, respectively. The mean importance rating for each skill/knowledge element included in the surveys is an average of respondents' individual importance ratings on a four-point scale, with 4 meaning very important, 3 meaning important, 2 meaning somewhat important, and 1 meaning not important. For analytical purposes, mean importance ratings of 2.50 and above identify skills and knowledge deemed important for incoming postsecondary students already to have mastered, while mean importance ratings of below 2.50 suggest skills and knowledge not important for incoming students already to have mastered. For the Standard English Conventions test section content domain, only the grand mean importance rating (the average of mean importance ratings across 25 skill/knowledge elements) is included due to the extensive number of testing points in this area.

**TABLE 8.1** KEY POSTSECONDARY CURRICULUM SURVEY FINDINGS: READING AND WRITING, ALL SURVEYED FACULTY

| RW Section Content Domain | Skill/Knowledge Element | Mean Importance Rating, Standard Deviation |
|---|---|---|
| **Information and Ideas** | Read closely to … | |
| | … identify information and ideas stated explicitly in a text | 3.77, 0.46 |
| | … draw reasonable inferences and conclusions from a text | 3.74, 0.48 |
| | Cite the textual evidence that best supports a given claim or point | 3.36, 0.83 |
| | Analyze data displays to … | |
| | … understand the information the graphic conveys | 3.27, 0.68 |
| | … synthesize information in the graphic with information conveyed in words | 3.22, 0.72 |
| | Determine central ideas in a text | 3.57, 0.65 |
| | Understand cause-effect, compare-contrast, and sequential relationships in text | 3.57, 0.62 |
| **Craft and Structure** | Determine the meaning of words and phrases used frequently in a wide range of academic texts (tier two words and phrases) | 2.99, 0.81 |
| | Analyze the purpose of part of a text or the text as a whole | 3.17, 0.83 |
| | Describe the overall structure of a text | 2.79, 0.90 |
| | Determine the point of view or perspective from which a text is related | 2.96, 0.88 |
| | Analyze the influence of point of view or perspective on a text's content or style | 2.77, 0.92 |
| | Synthesize information and ideas from multiple texts | 3.25, 0.86 |
| **Expression of Ideas** | Produce writing that … | |
| | … develops a logical argument by supporting a claim with cogent reasoning and relevant and sufficient evidence | 3.49, 0.73 |
| | … informs the reader or explains a concept, process, or the like | 3.40, 0.76 |
| | Create effective transitions (words, phrases, sentence) between and among information and ideas | 3.11, 0.85 |
| | Ensure precision of language for clarity and appropriateness to task, purpose, and audience | 3.40, 0.79 |
| | Use various sentence structures to achieve various rhetorical purposes (e.g., emphasis) | 2.75, 0.91 |
| **Standard English Conventions** | 25 skill/knowledge elements | 3.01, 0.73 (grand mean importance rating) |

Source: College Board (2019), pp. 38–41 (Appendix A.1), 44–46 (Appendix A.4)

**TABLE 8.2** KEY POSTSECONDARY CURRICULUM SURVEY FINDINGS: MATH, POSTSECONDARY MATH FACULTY

| Math Section Content Domain | Skill/Knowledge Element | Mean Importance Rating, Standard Deviation |
|---|---|---|
| **Algebra** | Represent contexts using a … | 3.67, 0.66 |
| | … linear expression or equation in one variable | 3.67, 0.66 |
| | … linear equation in two variables | 3.26, 0.91 |
| | … linear inequality in one or two variables | 2.82, 0.95 |
| | … system of linear equations | 2.93, 0.96 |
| | … system of linear inequalities | 2.39, 0.92 |
| | Interpret variables, constants, and/or terms in a linear equation | 3.73, 0.59 |
| | Evaluate a linear expression | 3.71, 0.61 |
| | Solve a … | |
| | … linear equation | 3.77, 0.54 |
| | … system of two linear equations | 2.96, 0.98 |
| | Graph a … | |
| | … linear equation | 3.71, 0.63 |
| | … linear inequality | 3.06, 0.95 |
| | … system of two linear equations | 2.86, 0.99 |
| | … system of two linear inequalities | 2.30, 1.00 |
| **Advanced Math** | Understand numbers and number systems, including radicals and exponents and their qualities | 3.48, 0.74 |
| | Represent contexts using a(n) … | |
| | … quadratic equation in two variables | 2.56, 1.02 |
| | … exponential equation in two variables | 2.27, 1.02 |
| | Interpret variables, constants, and/or terms in a(n) … | |
| | … quadratic equation | 3.33, 0.91 |
| | … exponential equation | 2.90, 0.99 |
| | Use properties of variables to … | |
| | … add, subtract, and multiply polynomials | 3.55, 0.78 |
| | … divide polynomials | 2.86, 0.96 |
| | … factor polynomials | 3.32, 0.93 |
| | Evaluate a(n) … | |
| | … polynomial expression | 3.45, 0.76 |
| | … rational or radical expression | 3.12, 0.99 |
| | … exponential expression | 3.04, 0.98 |
| | Solve a(n) … | |
| | … quadratic equation | 3.31, 0.91 |
| | … polynomial (degree three or higher) equation in one variable | 2.48, 0.96 |

| Math Section Content Domain | Skill/Knowledge Element | Mean Importance Rating, Standard Deviation |
|---|---|---|
| Advanced Math (continued) | … rational or radical equation in one variable | 2.79, 0.95 |
| | … system of one linear equation and one nonlinear equation | 2.28, 1.01 |
| | Choose and produce equivalent forms of a quadratic or exponential equation | 2.66, 0.94 |
| | Isolate one variable in terms of other variables of an equation | 3.59, 0.71 |
| | Graph a(n) … | |
| | … quadratic equation | 3.22, 0.98 |
| | … polynomial (degree three or higher) equation in one variable | 2.50, 0.99 |
| | … exponential equation in one variable | 2.63, 1.03 |
| | Use and interpret function notation | 3.34, 0.93 |
| **Problem-Solving and Data Analysis** | Understand numbers and number systems, including … | |
| | … absolute value of real numbers | 3.38, 0.75 |
| | … elementary number theory (primes, prime factorization, divisibility, number of divisors, odd/even) | 3.26, 0.87 |
| | Solve problems with rates, ratios, and percents | 3.49, 0.76 |
| | Use units and unit analysis to solve problems | 3.20, 0.88 |
| | Identify and distinguish linear and exponential growth | 2.74, 0.98 |
| | Given a scatterplot, model statistical data with a(n) … | |
| | … linear function | 2.65, 1.12 |
| | … quadratic function | 2.09, 1.02 |
| | … exponential function | 2.03, 0.97 |
| | Solve problems using … | |
| | … measures of center, including mean, median, and mode | 2.60, 1.04 |
| | … measures of spread, including range and standard deviation | 2.24, 1.05 |
| | … sample statistics and population parameters | 2.09, 1.05 |
| | … probability | 2.18, 1.03 |
| | Understand the characteristics of well-designed studies, including the role of randomization in surveys and experiments | 2.06, 1.01 |
| | Read and interpret statistical graphs | 2.48, 1.08 |
| **Geometry and Trigonometry** | Solve problems using … | |
| | … area and volume formulas | 3.35, 0.75 |
| | … special right triangles | 2.99, 1.01 |
| | … the Pythagorean theorem | 3.39, 0.92 |
| | … theorems of triangle similarity and congruence | 2.69, 1.05 |
| | … circle theorems | 2.52, 1.01 |
| | … logical reasoning and mathematical proofs | 2.62, 1.00 |
| | … trigonometry relationships, including sine, cosine, and tangent | 2.50, 1.19 |

Source: College Board (2019), pp. 58–59 (Appendix B.1), 61–62 (Appendix B.4)

*Note.* Mean importance ratings from postsecondary math faculty only were considered for this analysis, rather than ratings from the full respondent sample (as was used for English language arts/literacy) because the math skill/knowledge elements are most prerequisite for readiness for college-level math courses. Ratings from social science and science faculty to the math skills/knowledge elements listed above were generally lower, particularly in social science.

The data in Table 8.1 and Table 8.2 broadly confirm the validity of the selection of skills and knowledge being tested on the digital SAT Suite. In Table 8.1 (English language arts/literacy), all skill/knowledge elements (including the 25 elements in the Standard English Conventions content domain considered together) are rated at or above 2.50, the threshold at which an element was considered important in the analysis. In Table 8.2 (math), most skill/knowledge elements have a mean importance rating of 2.50 or higher, with the relatively small number that do not meet or exceed this threshold being included in the design for the sake of coherence and comprehensiveness of math domain testing.

In 2024 College Board will again undertake a national curriculum survey, this time tuned to the skills and knowledge tested in the digital suite.

## 8.2.2 Alignment to State College and Career Readiness Standards

### Internally Generated Alignments

Although the digital SAT Suite has not been designed to measure any particular set of state college and career readiness standards for K–12 students, initial alignment studies undertaken by College Board indicate strong alignment to these standards generally. This is no accident, as the digital SAT Suite and these standards documents are derived from the same sorts of evidence about essential prerequisites for college and career readiness. College Board will issue alignment reports for each set of state standards and will update these documents as states revise their expectations. These reports, focused on clarity and usability, summarize the degree of alignment in reading, writing, language, and math as well as provide detailed tables showing the match between specific standards and skill/knowledge elements tested on the digital SAT Suite and vice versa.

### Independent Third-Party Alignments

To supplement and validate the internally produced alignment studies discussed above, College Board has contracted with an independent, third-party firm with extensive experience in alignment work to evaluate the degree of match between the digital SAT and the Common Core State Standards (CCSS).

CCSS was chosen as the basis for this alignment study because of its wide use in the U.S. educational system and because, like the digital SAT Suite, CCSS is grounded in high-quality evidence regarding essential college and career readiness outcomes. The tests of the digital SAT Suite were not designed to measure CCSS outcomes specifically, and College Board's internally developed alignments (mentioned above) document strong alignment to academic standards in both Common Core and non–Common Core states. Moreover, College Board is assisting and will continue to assist states that wish to conduct independent, third-party alignments to their own specific sets of standards.

The digital SAT–to–CCSS alignment study, completed in 2023, reached the following conclusions:

1. The overall purposes, content domains, and emphases of the digital SAT correspond to the overall purposes, content domains, and emphases of the CCSS.
2. A content analysis of four digital SAT forms as well as a stratified sample of items from the digital-suite item pool demonstrated that:
   a. All Reading and Writing and Math items aligned to at least one CCSS standard.
   b. All CCSS content domains, with the intentional exclusion of Speaking and Listening, were represented by items in the digital-suite item pool.
   c. The digital-suite items demonstrated the full range of cognitive complexity expected by the CCSS.
3. The test forms overall were considered acceptably aligned to the CCSS standards.

Additional research studies for the PSAT-related assessments are underway and expected to be completed in 2024.

## 8.2.3 Subject Area Validity Evidence

In an ongoing fashion, College Board collects and assesses high-quality evidence about what matters most for college and career readiness in the subject areas sampled by the tests. This evidence helps inform both high-level design decisions as well as which particular skill/knowledge elements are assessed and how.

The following sections summarize the highlights of this evidence in both English language arts/literacy and math. These evidence précis have been consolidated from essay-length pieces commissioned by College Board and written by experts in the various topics addressed; following each précis is a brief discussion of the topic's links to the digital SAT Suite. Full treatments of the evidence, including instructional implications and classroom implementation advice, can be found in College Board's classroom practice guides for English language arts/literacy (**satsuite.collegeboard.org/media/pdf/sat-suite-classroom-practice-elal.pdf**) and math (**satsuite.collegeboard.org/media/pdf/sat-suite-classroom-practice-math.pdf**).

### English Language Arts/Literacy

Six topics are addressed in the following subsections: (1) text complexity, (2) close reading and evidence use, (3) inferences, (4) vocabulary and knowledge, (5) Standard English conventions, and (6) disciplinary literacy. Each subsection discusses why the given topic is important for college and career readiness for all students and how the topic is addressed on the digital SAT Suite tests.

### Text Complexity

#### Why Text Complexity Is Important

Every U.S. state and dominion now has college and career readiness standards requiring that students be given access to grade-appropriate complex texts, an emphasis that began with the Common Core State Standards as outlined by the National Governors Association (NGA) Center for Best Practices and the Council of Chief State School Officers (CCSSO) (NGA Center for Best Practices & CCSSO, 2010a).

While significant challenges remain to getting all students fluent with complex text no later than the end of high school, there is no question about the centrality of the ability to read and analyze such texts independently to college and career readiness. Studies have demonstrated that this ability is a crucial differentiator between those who are college ready and not college ready (ACT, 2006; Nelson et al., 2012).

### Text Complexity on the Digital SAT Suite

Text complexity is a key consideration on the digital SAT Suite Reading and Writing section. Each of the suite's testing programs presents test takers with broadly defined ranges of appropriately challenging texts across various subject areas. Table 8.3 summarizes the text complexity ranges sampled by each of the testing programs.

**TABLE 8.3 DIGITAL SAT SUITE READING AND WRITING TEXT COMPLEXITY RANGES BY TESTING PROGRAM**

| Digital SAT Suite Program | Text Complexity Range |
| --- | --- |
| SAT | Grades 6–8, 9–11, and 12–14 |
| PSAT/NMSQT and PSAT 10 | Grades 6–8, 9–11, and 12–14 |
| PSAT 8/9 | Grades 6–8 and 9–11 |

As the table indicates, grades 12–14 texts do not appear on PSAT 8/9, as College Board made the determination that these texts were too advanced to be appropriate for use in assessing eighth and ninth graders.

To ascertain a text's complexity, College Board uses both a robust quantitative measure and a qualitative rubric. The quantitative tool takes a text of any size and produces three measurements: Syntactic Complexity, Academic Vocabulary, and an overall model prediction. The Syntactic Complexity measure evaluates more than two dozen text attributes, including mean sentence length before the sentence root, the number of dependent clauses per sentence, and intersentence cohesion. The Academic Vocabulary measure evaluates more than a dozen text attributes, including the average frequency with which words in the text appear in a corpus of college-level textbooks, the average age at which people typically acquire the words in the text, and the average concreteness of words in the text. Syntactic Complexity and Academic Vocabulary are calculated values. The model prediction is inferred from a model that has been trained on the CommonLit dataset binned into the ranges used on the digital SAT Suite.

## Close Reading and Evidence Use

### Why Close Reading and Evidence Use Are Important

The driving question for K–12 educators about text complexity is what to do to provide access to complex texts for all their students, not just the students performing well enough to be in traditional college preparatory tracks in high school. Two research-based means of attaining such access are employing close reading techniques and making regular use of evidence. Using close reading techniques and identifying and discussing evidence are highly efficient means of attaining competencies in literacy closely linked to readiness for and success in college, workforce training, and civic engagement in a democratic republic. In particular, the ability to identify and deploy evidence when reading and writing analytically is consistently highly ranked in polls of employers and college faculty (Hart Research Associates, 2018; ACT, 2016, 2018, 2020; College Board, 2019; Intersegmental Committee of the Academic Senates of the California Community Colleges, the California State University, and the University of California, 2002). Facility with evidence is also considered essential to attaining the academic literacies that enable students from a variety of minority cultural and linguistic backgrounds to integrate successfully into postsecondary academic and technical settings (Preto-Bay, 2004; Papashane & Hlalele, 2014).

*Close reading* is sustained, purposeful intellectual work that centers on carefully reading a brief but rich and complex text (or excerpt from a longer work) in order to understand what the text says and how it says it (Beers & Probst, 2012; Fisher et al., 2014; Shanahan, n.d.; Lapp et al., 2015). *Evidence* is support within a text itself, in such forms as direct quotations, paraphrases, structural elements, and quantitative data, for a reader's interpretive claim regarding the text. Evidence from text is marshaled in support of an answer to a question—either the reader's own or one posed to the reader—regarding the information, ideas, or events the text is communicating.

Gathering such evidence is arguably the primary activity readers engage in when reading closely. All other reading-related activities—for example, monitoring comprehension, questioning the text, rereading, and summarizing while reading—circle back to evidence gathering. Students have to read closely in order to locate the evidence needed to answer their own and others' questions about what an author is saying, make an effective point in a discussion, or prepare a formal response to the text. Reading for evidence demands the careful attention that is the hallmark of close reading. In turn, the process of collecting evidence returns the reader, sometimes repeatedly, to the text in a focused way. Seeking evidence provides a purpose and structure for close reading and, in so doing, leads to more careful consideration of the text than does reading with a less-clear aim.

The motivation for developing and assessing students' command of textual evidence (e.g., quotations and paraphrases) can be boiled down to three key considerations:

- Evidence skills are highly valued across and beyond disciplines.
- Comprehension, analysis, and use of evidence differentiate experts from novices.
- Exposure to and practice with evidence in reading contexts improves reasoning.

The centrality of evidence skills to educational standards is difficult to overstate. Facility with evidence is a key standard in the National Assessment of Educational Progress (NAEP) Writing and Reading frameworks (National Assessment Governing Board, 2017, 2019), an anchor standard of the Common Core State Standards for Reading, for Writing, and for Speaking and Listening (NGA Center for Best Practices & CCSSO, 2010a), a major pillar of the Next Generation Science Standards (NGSS) (NGSS Lead States, 2013), and one of the four main dimensions of the *College, Career, and Civic Life (C3) Framework for Social Studies State Standards* (National Council for the Social Studies, 2013). College Board's 2019 National Curriculum Survey Report (College Board, 2019) shows that postsecondary faculty across disciplines place high value on students' skill at identifying relevant evidence. Beyond academic disciplines, scholars and public commentators routinely identify facility with evidence as a crucial component of civic life (e.g., Rosenfeld, 2019).

Research on the cognitive practices of disciplinary novices and disciplinary experts shows that a major differentiator between the two groups is how they comprehend and analyze evidence: low-skill and high-skill individuals differ not merely in what they know about a given subject but also in what they know about "how to establish warrant and determine the validity of competing truth claims" using evidence (Wineburg, 1991). Think-aloud studies show that, when reading, expert readers analyze and evaluate evidence presented in the text significantly more than do novice readers (Nelms & Segura-Totten, 2019). Linguistic evidence shows that when speaking or writing about target subjects, the biggest differentiator between experts and novices is that the former are much more likely to use words associated with causal relationships and cognitive processes and much less likely to use emotion-oriented language (Kim et al., 2011), hallmarks of evidence-based reasoning and communication. Experts are more likely to seek out evidential information and to deploy it than novices are (Peffer & Ramezani, 2019).

One of the more intriguing findings of recent years regarding evidence skills is that reading about experiments that provide evidence relevant to claims is associated with greater conclusion accuracy and generalization ability than is performing those experiments (Renken & Nuñez, 2009). Though this may seem counterintuitive, it accords with findings that suggest that print exposure to evidence-based arguments improves readers' own argumentation. Research suggests that direct instruction in argumentation coupled with analysis of arguments presented in text improves reasoning skills (e.g., Larson et al., 2009), that reading-based inquiry instruction models improve argumentation (Probosari et al., 2019), and that the most effective intervention in this regard is instruction in warrants that link evidence to claims (von der Mühlen et al., 2019). Data suggest that evidence and argumentation skills acquired this way can be applied to new and varied contexts (Zohar & Nemet, 2002). College Board's hope is that digital SAT Suite items assessing command of textual evidence will not merely provide an assessment of students' skills with evidence but will also actually improve those skills by providing focused practice in text-based evidential reasoning.

As with textual evidence, the motivation for developing and assessing students' command of quantitative evidence (i.e., data from informational graphics) can similarly be distilled succinctly:

- Graphical literacy is highly valued across and beyond disciplines.
- Exposure to and practice with quantitative evidence improves data literacy.

The importance of graphical literacy has been recognized for decades (e.g., Fry, 1981), and graphical literacy is a key component of the NAEP Writing and Reading frameworks (National Assessment Governing Board, 2017, 2019), the Common Core State Standards for Reading and for Writing (NGA Center for Best Practices & CCSSO, 2010a), the Next Generation Science Standards (NGSS Lead States, 2013), and the *College, Career, and Civic Life (C3) Framework for Social Studies State Standards* (NCSS, 2013). College Board's 2019 National Curriculum Survey Report (College Board, 2019) shows that postsecondary faculty across disciplines place high value on students' ability to read, understand, and analyze graphical data displays. Additionally, numerous scholars and commentators have noted that graphical literacy is an essential component of everyday life in the 21st century (e.g., Galesic & Garcia-Retamero, 2011).

Evidence suggests that high levels of exposure to visually presented information in daily life are insufficient to achieve graphical literacy: so-called digital natives do not inherently have significantly developed visual literacy skills (Brumberger, 2011). Instead, students' data literacy and facility with graphics improves through exposure to authentic data in learning contexts (e.g., Kjelvik & Schultheis, 2019). Explicit instruction in graphical literacy that makes use of complex, realistic data is associated with both greater student comprehension of graphically presented data and improved student attitudes toward working with graphics (Harsh & Schmitt-Harsh, 2016). This effect has been found among children (e.g., Phillips, 1997) and college students (e.g., Picone et al., 2007) and is strongest for relatively simple graphs, such as bar graphs (ibid.). Research suggests that students should practice working with both specific data points and inferred trends (Tairab & Khalaf Al-Naqbi, 2004). College Board again hopes that these items help improve students' graphical literacy by providing opportunities for deliberate, meaningful practice.

The careful attention that evidence collecting requires provides a payoff in the form of deepened comprehension. Whether pursuing their own learning goals or responding to questions or tasks presented to them, students need to pay careful attention to the text. The brain activates while reading, and the brains of successful readers activate in ways different than those of less proficient readers (Wolf, 2018). Collecting evidence is one means of forcing the kind of attention and careful reading that can achieve deep understanding.

## Close Reading and Evidence Use on the Digital SAT Suite

### Close Reading

The digital SAT Suite Reading and Writing section places a premium on close reading. All test items in the Information and Ideas content domain stress the comprehension and analysis of brief, rich texts—ideal for close reading—sampled from a range of academic disciplines and representing ways of reasoning and using evidence in those fields. Information and Ideas items ask test takers to, for example, determine which of four quotations from a work of literature (prose fiction, poetry, drama, or literary nonfiction) best supports an interpretive claim about a character, narrator, or speaker; make a reasonable, text-based inference about the significance of a scientific phenomenon; and accurately and reasonably use data from a table or graph to assess the outcome of a governmental policy. Central Ideas and Details items in this content domain ask students to determine the main points and understand the key supporting information of texts. As the name implies, Inferences items call on students to make reasonable, text-based inferences by using explicitly stated and implied information and ideas. Although Information and Ideas items draw heavily on the abilities of close reading, analysis, and reasoning, they do not require test takers to have deep background knowledge on the topics addressed or to have read any of the published works, such as novels or plays, that the items draw from. All the information needed to answer the items is provided as part of the items themselves.

Other types of items on the digital SAT Suite also call on close reading skills. In the Craft and Structure content domain, students must extend their close reading skills to address the two topically related passages found in Cross-Text Connections items and to assess the rhetorical effects of authorial choices in Text Structure and Purpose and in Words in Context items. In the Expression of Ideas content domain, Rhetorical Synthesis items require test takers to selectively integrate information and ideas provided in bulleted list form in order to achieve a specified writerly goal. In these items, test takers are presented with a series of two or more factual propositions situated in an academic discipline such as science or history/social studies and directed to strategically blend the content conveyed into a single, often syntactically sophisticated sentence. Each item specifies the particular rhetorical goal to be achieved, such as to support a generalization or to emphasize a contrast. The sentence test takers select from the four offered answer choices must not only include only relevant information (and exclude irrelevant information) from the set of propositions but must also achieve the specified goal. These items do not address the conventions of Standard English—all answer choices are grammatical—but rather focus on the combining and blending of information in the service of indicated writerly goals. (Test takers' mastery of core conventions of Standard English are addressed by other test items; see "Standard English Conventions," below.) Transitions items, in the Expression of Ideas content domain, require test takers to determine the most logical transition word or phrase between or among information and ideas presented in texts.

## Command of Evidence

In Command of Evidence: Textual items, test takers are presented with scenario-based assertions and must select the evidence that best supports, illustrates, or weakens the assertion, as directed by the item. Textual evidence items in informational contexts are typically based on published studies, research papers, and similar works. These items could describe a hypothesis, claim, or conclusion and ask students to identify the evidence that would strengthen or weaken that hypothesis, claim, or conclusion. Textual evidence items in literature contexts may present a claim about a significant work of U.S. or world literature and ask students to evaluate quotations from that work to determine which one best supports the provided claim. Alternatively, these items may describe a technique or pattern, such as an author's use of repetition, and ask students to identify which of the provided quotations from a work by that author best illustrates the author's use of that technique or pattern.

To ensure that informational textual evidence items assess students' ability to link evidence with assertions rather than students' knowledge of actual states of affairs, these items present evidence in conditional terms, asking students to identify the choice (such as the result of an experiment) that would be the best evidence for or against a proposition if that choice were true. This means that these items may present scenarios or evidence that are hypothetical; for example, an item may posit the existence of a follow-up study to evaluate an actual previous finding or may present evidence that could have been collected in an actual study but was not. It also means that students do not need to know or even consider whether a particular finding actually occurred to successfully answer the items. Similarly, textual evidence items in literature contexts do not require prior knowledge of the texts in question; the relevant assertion for which students are asked to find the best evidence is presented with the stimulus text, and the quotations in the answer choices are evaluable on their own, without knowledge of their context in the original work.

The scenarios and assertions in Command of Evidence: Textual items are also representative of situations and tasks students encounter in academic settings. For these items, scenarios are closely aligned with one of the several academic domains sampled by the digital SAT Suite Reading and Writing section (literature, history/social studies, the humanities, science) and their subdomains. For example, when the domain sampled is literature, a key driver of representativeness is the selection of literary works used in the items, which reflect mainstream literature curricula (e.g., *Ethan Frome* rather than *Smuggler's Run: A Han Solo & Chewbacca Adventure*). For all Command of Evidence: Textual items, an important consideration for representativeness is that the evidence presented reflect disciplinary standards and typical practices (e.g., personal experiences have little evidentiary power in the sciences; literary analysis distinguishes between assertions made by characters and the views of the author).

Since Command of Evidence: Textual items assess test takers' ability to reason about evidence, the evidence presented links to the assertion via inference. That evidence doesn't, in other words, simply restate or paraphrase the assertion or otherwise make the item a literal comprehension task. Naturally, the scope and challenge of reasoning required varies based on testing population, intended item difficulty, and the particular context being considered, but all items require at least some reasoning on test takers' part.

In Command of Evidence: Quantitative items, test takers are presented with brief texts accompanying figures (tables, line graphs, bar graphs). These texts may describe the figure with little reference to outside information, may describe the actual circumstances surrounding the data (e.g., conditions under which data were collected, a hypothesis being evaluated, relevant historical or scientific context), or may present a hypothetical scenario pertaining to the data. Texts are accurate (if describing real circumstances) or plausible (if describing hypothetical scenarios). When a Command of Evidence: Quantitative item includes a description of actual circumstances, such as a specific team's hypothesis or methodology, test developers consult the original research paper or other relevant documentation to ensure that the description is accurate. Note, however, that the description in the stimulus is not exhaustive: it includes only information that is necessary to clarify the data and support students' completion of the task. Both to avoid unnecessary time expenditure on the part of students and to keep the focus of Command of Evidence: Quantitative items on graphical literacy skills, informational "noise" in stimuli is minimized. The data, scenarios, and assertions in such items are not merely plausible but also representative of data, situations, and tasks students encounter in academic settings. This goal is achieved by closely aligning items and the data they use with one of the several academic domains sampled by the digital SAT Suite Reading and Writing section and their subdomains.

The levels of cognitive complexity and item difficulty in Command of Evidence: Quantitative items vary considerably, ranging from straightforward data-point identification through analyses of data patterns to sophisticated syntheses of data with information and ideas in stimulus texts. This range of approaches and challenge provides valuable information about students' graphical data literacy skills across a wide range of attainment.

## Drawing Inferences

### Why Drawing Inferences Is Important

Decades of research (e.g., van den Broek & Helder, 2017) have established that beyond fundamental decoding skills and vocabulary knowledge, much of what one thinks of as reading comprehension is, in fact, a web of inferencing skills. Not only sentence-by-sentence processing but also word-by-word processing appears to be highly influenced by causal inferencing on the part of readers (Kuperberg et al., 2011). Readers use inferencing skills both to impose coherence on texts that do not cohere (or are perceived to not cohere) and to determine the purposes and significance of textual elements and information in texts that are perceived as coherent (Graesser et al., 1994). So central is inferencing to reading that many comprehension failures are, in fact, inferencing failures (Cain & Oakhill, 1999).

*Inferencing* is a very broad category of activity, covering everything from simple mental substitutions of nouns for textual pronouns to drawing complex conclusions requiring multiple steps to reach. Early and simple inferencing ability is a strong predictor of later comprehension ability (Oakhill & Cain, 2012). As readers develop in skill and maturity, the number and complexity of inferences they can draw from a text increase (Casteel & Simpson, 1991), and as readers reach fluency, reading comprehension is strongly mediated by logical reasoning skills (Segers & Verhoeven, 2016), with inferential reasoning skills surpassing memory of the relevant text as a driver of comprehension (Oakhill, 1984).

Scholars have pointed out that while teaching logical inferencing is relatively straightforward, applying it (or general "critical thinking" skills) is much harder because inferencing typically must occur in knowledge domain contexts, to which people may struggle to map abstract inferencing skills (e.g., Willingham, 2019). Although there may be general inferencing skills, those skills are rarely activated in decontextualized settings (Kendeou, 2015). And not only are inferencing skills activated in contexts, but research also suggests that exposure to specific reading contexts—namely, complex texts with logically stated arguments—improves reasoning skills along with general comprehension (Osana et al., 2007). Assessing inferencing skills through context-based tasks is thus justifiable not only as a representative reading task but also because exposing students to texts with abundant causal and logical structures is productive of the very skills being assessed. In other words, exposing students to logically stated arguments both allows for the assessment of student inferencing ability and can improve student inferencing ability.

Given the centrality of inferencing skill to reading comprehension, it is unsurprising that the ability to make logical inferences from texts is included under the first standard in the Common Core ELA/Literacy College and Career Readiness Anchor Standards (NGA Center for Best Practices & CCSSO, 2010a) and was rated by postsecondary faculty as the second-most important reading skill (after explicit comprehension) in College Board's 2019 National Curriculum Survey Report (College Board, 2019).

### Drawing Inferences on the Digital SAT Suite

Inferencing, whether at the word, sentence, or textual level, is a routine requirement of the digital SAT Suite Reading and Writing section. The sort of complex inferencing described above, however, is most clearly represented in the Inferences items in the Information and Ideas content domain.

Consistent with the evidence suggesting the educational value of encountering reasoning-dense prose as well as with the digital SAT Suite's aim of offering an efficient testing experience, Inferences items distill the logical relationships inherent in longer pieces of complex, college-level prose down to single, tightly reasoned argumentative units. Each unit is presented up to the point at which the conclusion is introduced, and test takers must select the choice that most logically completes the argument. This approach allows the items to focus students' time and attention on only the material relevant to the skill being assessed while still exposing students to the kinds of logically dense texts that are characteristic of higher-level reading contexts and that help students improve as readers and thinkers.

The texts for Inferences items reflect the logical density of authentic academic or high-level general-interest prose. The texts do not take the structure of a formal logical proof but rather represent plausible approximations of argumentative units that a student would encounter in an academic reading context. Accordingly, Inferences texts are presented in prose that is clear, precise, and naturalistic (neither schematic on the one extreme nor "literary" on the other), and they build toward conclusions that are nontrivial for the target testing population.

Texts read like authentic writing in the various disciplines sampled by the Reading and Writing section, such as in the science-based sample below:

> Ecologists Anna Traveset and Nuria Riera investigated a decline in the population of the shrub *Daphne rodriguezii* from some areas of Spain's Balearic Islands. Traveset and Riera observed that the greatest population of *D. rodriguezii* is found in the area where Lilford's wall lizard (*Podarcis lilfordi*), which has been reduced in many parts of the islands, still thrives; that *P. lilfordi* appears to be the only natural disperser of *D. rodriguezii* seeds; and that seeds that are not consumed by *P. lilfordi* tend to accumulate beneath parent plants, where they are easily consumed by other animals and where they may struggle to thrive due to competition. Taken together, these observations suggest that ______

A. **the decline of the *P. lilfordi* population has contributed to the decline in the *D. rodriguezii* population**.

B. the decline in the population of *D. rodriguezii* may be attributable to an increase in the consumption of the plants' seeds by *P. lilfordi*.

C. potential dispersers of *D. rodriguezii* seeds have been outcompeted by *P. lilfordi*, leading to a decline in the population of *D. rodriguezii*.

D. the islands' population of *D. rodriguezii* must have been established before the island's population of *P. lilfordi*.

The text above stands in contrast to the hypothetical text below, which gets at the same general sort of inferencing skill but in a context rooted in logic rather than the discipline-based reasoning central to digital SAT Suite Inferences items:

> The perennial shrub *Daphne rodriguezii* is more widespread in Central America than is the flower species *Plumeria pudica*. *Daphne rodriguezii* is less widespread, however, than is the plant *Heliconia stricta*. It can be concluded that ______
> ***Heliconia stricta* is more widespread than is *Plumeria pudica*.**

The fact that Inferences items are written in naturalistic language and situated in real contexts means that there are likely to be multiple ways of expressing the conclusion of any given argumentative unit, and, in some cases, there may even be other conclusions that one could draw from the information but that are not presented in the answer choices and that are not as good as the keyed response in the sense that they are less likely, less complete, or less significant. An Inferences item, therefore, does not ask students to identify the correct form of the only valid conclusion, as in a formal proof, but rather the most logical conclusion among the answer choices given. The conclusion in the key follows from the information in the stimulus text and is a strong "real-world" conclusion, which allows test takers to select the key affirmatively rather than select the "least bad" option through elimination.

Although Inferences items are situated within recognizable academic domains, skilled readers are able to correctly answer the items with only the information provided in the items themselves and the general domain knowledge imparted by a typical rigorous high school curriculum. Inferences items may, for example, presume that students are familiar with the idea that organisms evolve but would not presume that students know what kinds of alterations in the nucleotide sequence of the genome are most likely to produce phenotypic changes. In other words, the assessment focus of Inferences items is on students' ability to draw reasonable inferences from complex texts, not on deep prior knowledge of the material presented.

## Vocabulary and Knowledge

### Why Vocabulary and Knowledge Are Important

The roles of vocabulary and knowledge in students' reading comprehension have long been overlooked in practice despite extensive research attesting to their importance. Instructional focus has instead been on the teaching and learning of discrete skills and strategies, often out of context, with the unrequited hope that they would transfer from one text to the next (Wexler, 2019). Skills and strategies do indeed have a role to play in increasing students' reading comprehension, but their value pales in comparison to that of vocabulary and knowledge.

Failure to understand and act on this fact renders many students unprepared for college and workforce training as they depart high school. Word and domain knowledge (as well as world knowledge) are essential to proficient reading comprehension, increasingly so as texts become more complex in higher grades. The failure to address this situation is one of the primary causes of the continuing gap in performance between struggling readers and their classmates who are able to access readings at or above their grade level.

The relationship between vocabulary and reading comprehension has been understood for nearly a century (Whipple, 1925; Chall & Jacobs, 2003). Decades of subsequent research have affirmed a close connection between vocabulary knowledge and reading comprehension skills (see, for example, Nation, 2009 for an overview). This association has been found in beginning readers (e.g., Silva & Cain, 2015), elementary school students (e.g., Quinn et al., 2015), middle school students (e.g., Lawrence et al., 2019), secondary school students (e.g., Ahmed et al., 2016), students with disabilities (e.g., O'Connor, 2014), second-language learners (e.g., Masrai, 2019), and readers of nonalphabetic languages (e.g., Dong et al., 2020).

In 2002, Isabel Beck, Margaret McKeown, and Linda Kucan introduced the notion of dividing up all words and phrases in English into three tiers as a way to create priorities within vocabulary instruction. In this scheme (Beck et al., 2013), tier one words and phrases (e.g., *family*, *fun*, *games*, *table*) are basic vocabulary and are commonly learned by children in their native language(s) through everyday discourse. Though young students will not necessarily learn all tier one words and phrases in English at the same rate, especially if they are learning English as a second or subsequent language, they will learn almost all of them sooner or later. Tier three words and phrases (e.g., *membrane*, *perimeter*, *manifest destiny*, *checks and balances*, *metaphor*) are used less frequently, and seldom in everyday conversation, and are generally specific to particular domains of knowledge (e.g., biology, geometry). Thus, they tend to appear in texts of only certain subjects, such as *tectonic* in geology texts (though tier three words and phrases sometimes "jump domains," as in "The election results signaled a tectonic shift in voter attitudes").

Tier two words and phrases (e.g., *influence*, *produce*, *variety*, *exclusive*, *particular*) are likely to appear in a wider variety of texts than are tier three words and phrases and, unlike their tier one counterparts, appear with increasing frequency the more sophisticated that text gets. Tier two words and phrases do not have a home in any one academic subject since they occupy texts universally. While subject-area teachers are eager to teach the tier three words and phrases that are the province of their disciplines (since these words and phrases often name the concepts in their fields) and while tier one words and phrases tend to be acquired through everyday discourse, tier two words and phrases are in danger of being left unattended, the responsibility of no one. Before the advent of college and career readiness standards, which shone a spotlight on the centrality of vocabulary and called out the special place of tier two ("general academic") vocabulary in students' K–12 and post–high school success, teachers tended to assume their students already understood the meaning of words and phrases in this category. If teachers thought about tier two words and phrases at all, they probably underestimated the frequency with which such vocabulary appears in the texts they assigned and failed to grasp the disproportionate role these words and phrases have in conveying texts' meaning (Snow, 2010; Adams, 2009).

Domain and world knowledge, too, support comprehension in a variety of ways (Britton & Graesser, 2014). Knowledge strengthens readers' ability to generate the inferences from text that lead to high-level comprehension, enhances readers' ability to combine information from parts of a text (or multiple texts) into a coherent understanding, and allows readers to integrate textual information with their prior knowledge.

## Vocabulary and Knowledge on the Digital SAT Suite

### *Vocabulary*

Although vocabulary knowledge and skills are so central to reading comprehension that any robust assessment of the latter will, to some degree, assess the former, there is merit to assessing vocabulary skills in a focused way. While vocabulary knowledge may be necessary for comprehension, it is not sufficient for comprehension (Biemiller, 2005). A variety of factors affect overall comprehension, making it difficult to provide useful information about student achievement, progress, and needs with regard to vocabulary acquisition and mastery if vocabulary skills are only assessed indirectly through more general comprehension tasks.

Adult (including college-level) readers rarely activate vocabulary knowledge in the absence of context. Instead, readers access and apply their vocabulary stores as they encounter or produce words in particular contexts (e.g., reading a news article or chemistry textbook, writing a sales report or term paper), and despite the renewed emphasis on direct vocabulary instruction in the 21st century, most vocabulary acquisition occurs through repeated contextual exposure (Stahl, 2003). Assessments of vocabulary knowledge and skill that do not reflect the contextual nature of vocabulary acquisition and activation may risk underestimating what students know and can do (Pearson et al., 2007). The importance of contextualized vocabulary is reflected in the findings of College Board's 2019 National Curriculum Survey Report (College Board, 2019), where postsecondary faculty rated the context-based understanding of word meanings as a skill of high importance.

In the context of the digital SAT Suite, the three-tier vocabulary model (Beck et al., 2013) is best thought of as a general framework for evaluating words under consideration for inclusion in test items. Words in Context items in the Craft and Structure content domain focus on high-utility academic, or tier two, words. These are words, typically acquired through direct instruction or reading exposure, that have broad applicability across academic and career contexts and are central to unlocking the meaning of text, particularly of the kinds of complex texts encountered in secondary and postsecondary instruction as well as in the workplace. It is worth recalling that word tiers do not (and are not intended to) include static, universally recognized banks of words. Words and senses of words rise and decline in frequency and general familiarity (e.g., the increasing prevalence of terms such as *application* and *program* in reference to computer technology), common words can have specialized meanings in different contexts (e.g., *selection* in a biology context vs. common speech), and people can reasonably disagree about whether particular words properly belong to one tier or another.

Additionally, there are many words that are not discipline specific (and so may not commonly be thought of as tier three) but are nevertheless sufficiently rare in writing across domains that they are not worth including in Words in Context items. Words such as *lassitude*, *supercilious*, and *adumbrate*—while they have real-world value and appeal in certain limited contexts—have low enough frequency to preclude their being tested profitably on the digital SAT Suite and are not the focus of Words in Context items.

In addition to the guidance of the three-tier framework, College Board uses empirical evidence to guide the selection of focal words for Words in Context items, namely:

- Test-based age-of-acquisition ratings—originally generated by Dale and O'Rourke (1981), updated and validated by Brysbaert and Biemiller (2017)—which provide test-based data about the age by which people tend to have acquired the meanings of given words
- Human-rater age-of-acquisition ratings (Kuperman et al., 2012), which indicate the age by which people believe that they had learned given words
- Word frequency data gathered by College Board from a corpus of hundreds of the most frequently assigned introductory-level college texts across subject areas, which reveal the words that students most need to know to comprehend the texts they are actually being assigned in U.S. colleges and universities

These measures, used in conjunction with the judgment of experienced test developers, ensure that the words being tested in Words in Context items are of high utility for college- and career-ready readers.

Words in Context items present high-utility academic words in rich contexts, by which is meant contexts reflective of college- and career-ready reading experiences and aligned with specified knowledge domains. Many Words in Context items discuss real people, places, research findings, books, artworks, events, and so on, and even Words in Context items that include generic elements are grounded in real contexts (e.g., an art critic's claim about abstract expressionist painters, or a researcher's study of panda metabolism).

***Knowledge***

Although the digital SAT Suite Reading and Writing section is expressly not a measure of test takers' knowledge in the subject areas the section samples, and while digital SAT Suite Reading and Writing items contain all the information needed to answer them correctly, they do call on test takers' abilities to read and comprehend appropriately challenging texts in these areas, to use critical reasoning and analytical skills developed in and particular to subject area courses, and to apply these skills and knowledge to items grounded in texts and contexts reflecting the academic demands of these areas. In other words, knowledge building in the subject areas lays the foundation for success on the tests and, more importantly, in test takers' postsecondary educational pursuits. For additional information on the discipline-based nature of the digital SAT Suite Reading and Writing section, see "Disciplinary Literacy," below.

## Standard English Conventions

### Why Standard English Conventions Are Important

Standard English is the variety of English that has tended to be most valued in academic and professional settings (Beason, 2001; O'Neill, 2018). Although there is some variation in the grammatical forms (such as passive voice) and levels of formality preferred in different academic disciplines and workplace settings, decades of research have shown that effective use of Standard English is a fundamental expectation in academic and professional settings. The term *Standard English* (also sometimes *Standardized English*) refers to the spoken and written language varieties that are expected in most institutional contexts in the United States, such as government and schools. The conventions of Standard English are the patterns, or "rules," of grammar, usage, punctuation, capitalization, and spelling that are generally accepted in the present day.

However, the conventions of Standard English are not just about rules and "correctness." They also contribute to clear and effective communication in academic and other institutional contexts. For instance, in Joseph Williams and Joseph Bizup's well-known book on writing, *Style: Lessons in Clarity and Grace* (Williams & Bizup, 2017), readers are taught to put their most important ideas and "actors" in the subjects of their sentences and to vary sentence length using subordinate clauses for rhetorical effect. Having a language to talk about grammatical concepts such as these can help students become aware of the conventions of Standard English in different disciplines and make deliberate, well-informed choices about how to use language for clear and effective written and formal spoken communication. Thus, understanding and controlling for the conventions of Standard English to accomplish specific purposes and to reach intended audiences are valuable academic and professional skills that contribute to college and career readiness.

Terms such as *conventions*, *usage*, and *effective communication* can help teachers convey the changing nature of Standard English more accurately than can terms such as *proper English*, *correct English*, and *rules*. *Conventions* and *usage* also reflect a descriptive view of Standard English rather than a prescriptive one. Prescriptive views of language are based in a static view of English as having just one "correct" variety and as being governed by a prescribed set of rules—even when those rules are rarely adhered to in practice. One example of a prescriptive rule is "Do not split an infinitive"—a directive that is regularly broken in written Standard English and whose violation is rarely viewed by readers as an error (Beason, 2001). Descriptive views of language, on the other hand, acknowledge that what counts as acceptable or effective Standard English changes over time and is determined by how real people use and respond to language patterns. Thus, descriptive views of Standard English seek to convey current uses of and perspectives on language conventions rather than a static and potentially outdated vision of what the conventions of Standard English "should" be.

In discussions of grammar and conventions, it is also helpful to distinguish Standard English from *vernacular* or *nonstandard dialects*. All languages, including English, encompass multiple varieties, or *dialects*. The term *dialect* refers to the patterns of language used by a particular group with a shared regional or social affiliation. Everyone speaks a dialect even if they are unaware of it. The terms *vernacular dialect* and *nonstandard dialect* help distinguish other language varieties from Standard English, the variety typically used and expected in academic and professional settings, but the use of those terms should not be taken to imply that these language varieties are less grammatical, less logical, or less communicative or expressive than Standard English. Some well-researched vernacular dialects in the United States include Appalachian English, African American English, and Chicano English. Decades of research have shown that valuing, discussing, and building on students' home languages and dialects benefit their language and literacy learning (Heath, 1983; Lee, 2007). Conversely, telling students that the nonstandard varieties of language they are using are wrong or improper can hinder students' language and literacy learning.

Teachers as well as students benefit from viewing the conventions of Standard English as tools for clear and effective communication in academic and professional settings rather than simply as rules. This descriptive, communicative perspective on Standard English changes the teacher's role from being a judge of whether prescriptive rules of grammar have been followed to being a coinvestigator of patterns of conventions and usage in different academic subjects and genres. It also provides teachers with a more productive answer to the student question "Why do we have to know this?" Developing students' awareness and command of the conventions of Standard English is beneficial to their future academic and professional pursuits, and this work can be undertaken in creative, engaging ways. By teaching conventions of Standard English as meaningful and useful, educators can empower students to succeed in college, the workplace, and beyond.

### Standard English Conventions on the Digital SAT Suite

The digital SAT Suite Reading and Writing section addresses core conventions of Standard English sentence structure, usage, and punctuation in context-bound ways focused on enhancing the communicative power of text rather than simple demonstrations of "correctness." Test items in the Standard English Conventions content domain take two main forms. Boundaries items require test takers to apply Standard English conventions when editing short texts (typically one or two sentences in length) to ensure that the resultant sentence(s) conventionally separate or join phrases, clauses, or sentences. These items address such matters as standard end punctuation, semicolon and colon use, and the conventional use (or nonuse) of punctuation, such as commas, dashes, and parentheses, to set off (or not set off) information and ideas within sentences. Form, Structure, and Sense items, on the other hand, assess test takers' ability to apply core Standard English grammar and usage conventions in context, such as ensuring subject-verb agreement, using verb tenses and aspects appropriately, and appropriately forming and using plurals and genitives (possessives). In both broad types of items, test takers work in authentic, meaningful contexts grounded in the academic disciplines, and the focus is on enhancing the communicative power and clarity of text.

## Disciplinary Literacy

### Why Disciplinary Literacy is Important

As students advance through school, the texts they read become more specialized. A second grader's social studies textbook is different from a high school junior's history book, and young children's science texts are akin to their social studies books in a way not true of high school texts in the same subjects. To read these more specialized texts properly—in ways that would lead to thorough comprehension and sophisticated interpretations appropriate to those disciplines—students need to approach them with a knowledge of a discipline and its purposes, content, and methodologies.

The term *content knowledge* refers to an awareness or understanding of information on a particular topic. Knowing the distinction between *meiosis* and *mitosis*, that the Great Depression began in 1929, and that *Narrative of the Life of Frederick Douglass, an American Slave* was one of three autobiographies written by this magisterial author, orator, and activist are all examples of content knowledge. It is important that students learn some of the facts and information (content knowledge) produced by the disciplines. However, other kinds of knowledge matter too.

Students should also develop knowledge of a discipline. This *disciplinary knowledge* encompasses an awareness of a discipline's purposes and methodologies: how and why experts do their work, what constitutes a reasonable claim, and how one can appropriately refute such claims. In a history class, it may be important that students learn what the Battle of the Bulge was (a German offensive during World War II) and some facts about it (e.g., the Germans were defeated). But disciplinary knowledge leads students to search for the causes of the battle, to ask why it was considered so significant, or to question the particular interpretation of it in the text they are reading. Students need to gain both content knowledge and disciplinary knowledge; they need to know not only the *whats* but also the *whys* and *hows* of a discipline.

It is this disciplinary knowledge that underlies a discipline's literate practices, and students must have such knowledge if they are to read and write appropriately within a discipline. Disciplinary knowledge includes an understanding of how a field creates, communicates, and evaluates information. Knowing about the discipline can help students understand whether a given text is important and, if it is, what in it is essential. Often students asked to highlight the important information in a text—a popular content area reading strategy—end up underlining nothing or everything because they lack the disciplinary insights that would allow them to distinguish the vital from the incidental (Dunlosky et al., 2013).

Students who recognize what is important in a history text (e.g., who the author is, historical figures' intentions) or science text (e.g., what processes are involved in mitosis or chemical reactions) are better able than their peers to separate wheat from chaff. Disciplinary awareness can help students identify and evaluate the evidence in written arguments. Experimental evidence, for instance, is especially important in arguments in science but not so much in history. Students can use knowledge of a discipline to determine the voice to adopt in writing, how to use the technical vocabulary of a field, and so on in ways consistent with the core beliefs, values, and practices in that field. Accordingly, literacy instruction with disciplinary texts should be closely aligned with the mores, normative standards, traditions, skills, and social discourse practices of the disciplines.

As different as the various disciplines and their specializations may be, one thing remains the same: experts in all fields read and write. Experts in scientific and other technical fields, for example, spend substantial amounts of time reading and writing (Kwon, 2017; National Science Foundation, 1976; Tenopir et al., 2004). Scientists read journal articles, review research literature, make grant applications, collaborate through email exchanges, create detailed records of experiments in laboratory notebooks, write journal articles and research reports, and engage in dozens of other daily reading and writing tasks in their work routines. It is fair to say that one could not participate in science successfully without the ability to read well and with great stamina and to communicate in writing in ways characteristic of science. Given the ubiquity of reading and writing within the disciplines, it seems only right that schools not only have students read and write throughout the curriculum but also give them explicit guidance in the special text features and ways of reading and writing specific to various fields of study.

One reason students struggle in college, the workplace, or the military is lack of sufficient literacy skills. Because so many students are underprepared, a high percentage of them require remediation in college, with about 40% of first-year postsecondary students nationwide requiring remedial support in reading or writing (Bautsch, 2013). The National Assessment of Educational Progress (NAEP) reports that only 37% of 12th graders taking the 2019 NAEP Reading assessment scored at or above the proficient level in reading (National Center for Education Statistics, n.d.). Especially worrying is that proficiency in literacy in the United States is highly unequal: according to 2018 data from the Programme for International Student Assessment (PISA), the gap in reading scores between students in the top and bottom quarters of the economic, social, and cultural status index in the United States was larger than that in all but two countries where it was measured (Schleicher, 2019).

According to the NAEP, the problem is not one of basic literacy. Nearly all students in the United States are able to read and write: they can sign their names, decode and understand simple messages, and the like. What is missing is the ability to read complex texts in sophisticated ways and to communicate complicated ideas subtly and persuasively—outcomes more likely to be accomplished through a disciplinary literacy approach than one aimed at trying to teach general reading comprehension or writing skills.

### Disciplinary Literacy on the Digital SAT Suite

The requirements of literacy in the disciplines deeply inform the digital SAT Suite Reading and Writing section. Texts appearing in the section reflect the demands of literacy in the disciplines of literature, history/social studies, the humanities, and science. Science and social science texts, for example, discuss hypotheses, methodology, data, conclusions, and implications and may be accompanied by informational graphics (tables and graphs) that display associated data and otherwise complement the information and ideas conveyed in words. Items throughout the Reading and Writing section call on test takers to respond in ways appropriate to the various disciplines. These demands begin with the stimulus texts associated with individual items. These texts, though brief, are richly reflective of the concerns, methods, and ways of thinking and creating knowledge in the various academic disciplines from which they sample. They pose scenarios, present information and ideas, assert claims, and offer evidence in ways that embody the norms and conventions of the subject areas. While literature items ask test takers to support interpretive claims about published works using actual quotations from the texts or present excerpts from published works for test takers to analyze, items set in science or social science contexts may ask test takers to accurately and skillfully use data from experiments or observational studies, represented in a table or graph, to support or challenge an argumentative claim appropriate to those fields. To answer these items successfully, test takers need not only broad-based reading comprehension and data analysis skills but also an understanding of how and for what purposes various subject areas create and convey knowledge. In this way (and in others), items on the Reading and Writing section encourage test takers' development of disciplinary knowledge in authentic ways, even as the items themselves provide all the information necessary to answer them without topic-specific background knowledge.

## Math

Four topics, each corresponding to one of the digital SAT Suite Math section content domains, are discussed in the following subsections: (1) algebra, (2) advanced math, (3) problem-solving and data analysis, and (4) geometry and trigonometry. Each subsection discusses why the given topic is important for college and career readiness for all students and how the topic is addressed on the digital SAT Suite tests.

### Algebra

#### *Why Algebra Is Important*

There has been substantial and sustained interest in promoting students' success in algebra for a generation or more. In large-scale studies (e.g., Adelman, 2006; Gamoran & Hannigan, 2000; Lee & Mao, 2021; Trusty & Niles, 2004), success in algebra has been linked to increased secondary and postsecondary course taking, improved high school and college graduation rates, and more productive job and career outcomes. For instance, in examining nationally representative longitudinal data to study the long-term educational and career trajectories of students who were enrolled in 10th grade in 1980 (i.e., presumptive members of the high school graduating class of 1982), Rose and Betts (2001) found that "math curriculum is strongly related to student outcomes more than 10 years later," including college graduation rates and earnings (p. xix). Notably, Rose and Betts found that "the biggest difference [among student outcomes] is between courses at or above the algebra/geometry level and courses below the algebra/geometry level" (pp. xix–xx), by which they meant vocational math and prealgebra.

Recognition of the importance of strong skills in algebra has led to efforts to incorporate algebraic thinking into the elementary school curriculum (e.g., Kieran et al., 2016). Furthermore, while taking first-year algebra (Algebra I) was historically a part of the ninth-grade curriculum, many students now take this course in eighth grade or earlier (Stein et al., 2011). In almost all districts, a passing grade in Algebra I is a requirement for high school graduation.

Facility with algebra opens many doors for students; lack of such facility carries the significant risk of keeping those doors shut, whether one considers educational or vocational aspirations. Mastering concepts taught in algebra courses is viewed as a key prerequisite on the path to higher-level math courses, particularly calculus (e.g., Kaput, 1995; National Mathematics Advisory Panel, 2008; Rakes et al., 2010; Stein et al., 2011). Biag and Williams (2014) extend this value proposition further by noting that students failing Algebra I (and potentially having to retake it) are in danger of being cut off from advanced high school science coursework given those courses' algebra prerequisite. Lack of algebra skills and knowledge can also inhibit or exclude students from pursuing a range of well-paying blue- and white-collar jobs, including, among others, careers as carpenters, electricians, millwrights, and sheet metal workers as well as actuaries, architects, dietitians and nutritionists, and market research analysts (Weedmark, 2018).

In short, lack of access to or success in high-quality Algebra I instruction is a significant and, arguably, insuperable barrier to students' academic success in high school, postsecondary education, and well-paying careers. Algebra, and specifically the Algebra I course, is, therefore, an enormously important milestone in students' math learning.

#### *Algebra on the Digital SAT Suite*

Coursework in algebra is very important in each student's math journey, and facility with algebra provides students with opportunity for further success, while lack of facility burdens them with risk of reduced opportunity. As a result, skills in algebra have significant representation on the digital SAT Suite exams. Items in the Algebra content domain of each exam align most closely with topics covered in a typical rigorous first-year secondary algebra course, including assessing the skills and knowledge associated with working with linear expressions, linear equations in one and two variables, linear functions, systems of linear equations, and linear inequalities. Test items cover such skills and knowledge as creating and using a linear equation; identifying an expression or equation that represents a situation; interpreting parts of a linear equation in context; making connections between linear equations, graphs, tables, and contexts; determining the number of solutions and the conditions that lead to different numbers of solutions; and calculating and solving. The test items aligned to algebra skill/knowledge elements range in difficulty from relatively easy to relatively complex and challenging. The test items require students to demonstrate skill in generalization, abstraction, and symbolization, with a strong emphasis on equivalence and using structure. Many of the test items are constructed to allow for more than one solving strategy.

## Advanced Math

### *Why Advanced Math Is Important*

This subsection proposes a working definition of *advanced math* focused on a broad conceptual divide between a key focus of Algebra I and higher-level math. While Algebra I attends centrally to the concept of linear equations and functions, advanced math, as treated here, focuses centrally on nonlinear equations and functions. One way to quickly conceptualize this distinction is to note that linear equations and functions graph as straight lines, while nonlinear equations and functions do not. Because nonlinear equations and functions are more conceptually complex than linear ones and because an understanding of the former builds on an understanding of the latter, nonlinear properties can reasonably be categorized as "advanced." Indeed, as students progress in their study of math, they build on their earlier experiences with algebraic expressions and linear functions to investigate the ways in which nonlinear equations and linear functions are powerful tools for making sense of and modeling phenomena in their worlds.

Advanced math skills and knowledge, as defined here, are relevant to secondary-level students in numerous ways. First, advanced math in high school serves as a bridge to still more advanced coursework in math in both high school and college, and it opens access to coursework in secondary and postsecondary science that has advanced math prerequisites. Carnevale and Fasules (2021), for example, pulled together data from the U.S. Census Bureau and the Occupational Information Network, a database sponsored by the U.S. Department of Labor's Employment and Training Administration, and found that "jobs in science, technology, engineering, and math (STEM) use the highest levels of math, with 92 percent of STEM workers needing to know at least Algebra II. … Most STEM jobs require even higher-level math, with 67 percent requiring college-level math such as calculus" (Carnevale & Fasules, 2021, p. 1). Thus, the study of advanced math is an essential pathway toward STEM-related professions.

Second, attaining advanced math skills and knowledge in high school is important for college and career readiness for students seeking entry into a wide range of blue- and white-collar occupations, both outside and, especially, within STEM fields. Based on statistical analysis of employment data as well as input from business leaders and over three hundred two- and four-year faculty, the American Diploma Project (2004) found a convergence between the knowledge and skills employers seek in new workers and those that college faculty expect of entering students. In particular, both employers and college faculty expect high school graduates to be able to apply math concepts typically taught in advanced secondary coursework in algebra. This finding for the continued work in math beyond a first course in algebra is consistent with the more recent recommendation from the report *Catalyzing Change in High School Mathematics: Initiating Critical Conversations* (National Council of Teachers of Mathematics, 2018). This report recommends that high schools offer continuous four-year math pathways, including two to three years in a common pathway that includes focused attention on learning the concept of function (one of the "Essential Concepts" in high school math).

Third, acquisition of advanced math skills and knowledge is associated with positive educational and economic outcomes for students. Advanced math is typically a requirement for entry into a four-year college. For example, in an analysis of college-going California high school students, Asim, Kurlaender, and Reed (2019) found that compared to the overall population of high school seniors, a significantly larger proportion of students who applied and were admitted to either a California State University or a University of California institution took advanced math courses (for which advanced algebra is a prerequisite) in their senior year. This is consistent with prior research that found similar correlations to college entry as well as to college completion (e.g., Gottfried et al., 2014; Long et al., 2012). Research also has identified correlations between higher earnings and completion of more and higher levels of math (e.g., Rose & Betts, 2004). Indeed, Moses and Cobb Jr. (2001) refer to algebra as the new civil right, as students who do not have access to higher-level math have less access to economic mobility.

Fourth, principles and methods of advanced math can be applied productively to analyze and understand a gamut of academic and real-world scenarios that students will encounter throughout their lives. Because many authentic applications, both within the field of math and in the real world, are nonlinear, students will need to work with quadratic, polynomial, rational, exponential, and other nonlinear functions. For example, quadratic functions are useful models for understanding and analyzing real-world situations such as forecasting business profit and loss, modeling projectile motion, and describing the movement of bouncing objects. Polynomial functions can be used to model the curves in a roller coaster, the concentration of a particular drug in the bloodstream, and other real-world situations. Rational functions are useful for analyzing real-world phenomena such as density, work, rates of change, and volume. Modeling with exponential functions is also important in such contexts as bacteria growth, the depreciating value of a vehicle, or the value of an investment over time. The study of these different nonlinear function types can develop both the habits of mind and habits of interaction that students need to become powerful users of math, to better interpret and understand their worlds, and to make better predictions about phenomena of interest.

#### *Advanced Math on the Digital SAT Suite*

The advanced math topics assessed on the digital SAT Suite exams extend those covered in the Algebra content domain topics into nonlinear equations and functions and align most closely with topics mastered in a typical rigorous second-year secondary algebra course and sometimes beyond. Since these Advanced Math test items build on skills and knowledge first mastered with linear expressions and equations, it follows that these topics should also be well represented on college and career readiness exams such as those of the digital SAT Suite. As a result, skill/knowledge elements in Advanced Math are represented on the digital SAT Suite exams in relatively high proportions.

The Advanced Math content domain assesses skills and knowledge associated with working with quadratic, exponential, polynomial, rational, radical, absolute value, and conic section equations and functions. Similar to Algebra items, test items in the Advanced Math domain cover such skill/knowledge elements as creating and using a nonlinear equation; identifying an expression or equation that represents a situation; interpreting parts of an equation in context; making connections between equations, graphs, tables, and contexts; determining the number of solutions and the conditions that lead to different numbers of solutions; and evaluating and solving using nonlinear equations and systems that include a nonlinear equation. The test items in the Advanced Math domain range in difficulty from relatively easy to relatively complex and challenging. Many of the test items represent challenging, authentic problems in context for which students can draw on strategies developed during their coursework to solve.

## Problem-Solving and Data Analysis

### *Why Problem-Solving and Data Analysis Are Important*

Data are everywhere, and working with, understanding, and learning from data have become necessities of daily live. Personal data are commonly collected through digital devices, and daily behaviors are routinely recorded. Businesses, governments, and other entities use data analytics and powerful computing technology applied to massive pools of information to inform decision making (Pence, 2014), with examples including developing marketing targeted to consumer interests; predicting rises and falls of demand for products and services; improving app-based navigation; aiding healthcare providers in suggesting courses of treatment; detecting financial fraud; and tracking the spread of foodborne illness (Rice, 2023; Helms, 2015).

Now more than ever, it is essential that all students leave secondary school prepared to live and work in a data-driven world (Engel, 2017). The development of statistical thinking and data acumen is imperative today, as every individual must use data to make informed decisions involving numerous aspects of their lives (National Academies of Sciences, Engineering, & Medicine, 2018; Wilkerson, 2020). Many college majors require coursework in statistics (American Statistical Association, n.d.), and statistician jobs are expected to grow by about 35% between 2020 and 2030 (U.S. Bureau of Labor Statistics, 2021). Postsecondary education in statistics is also changing to meet the demands of 21st-century life and careers, with the *Guidelines for Assessment and Instruction in Statistical Education (GAISE) College Report* calling for the preparation of students in statistics at the college level to shift from centering on the application of a list of formulas to a focus on developing the skills of interpretation and understanding data (GAISE College Report ASA Revision Committee, 2016). Enrollment of college students in statistics has steadily increased, with the latest (2015) data from the ongoing survey conducted by the Conference Board of the Mathematical Sciences (CBMS) of math and statistics departments at two- and four-year colleges and universities showing that 737,000 students took statistics courses as part of their undergraduate work (Blair et al., 2018). This represents a 56% increase in enrollment in statistics classes since the year 2000, the initial year the CBMS survey was administered.

With ubiquity of data comes responsibility. Although not all students will become statisticians or professional data analysts, they still must be able to check data sources and “mind” the data they encounter. Data minding (Meng, 2021) is a “stringent quality inspection process that scrutinizes data conceptualization, data pre-processing, data curation and data provenance” (p. 1161). In other words, students, regardless of their educational plans and intended career paths, must be data literate, able and disposed to act as knowledgeable users of data themselves as well as informed consumers of other people’s efforts to use data to support claims and guide actions.

### *Problem-Solving and Data Analysis on the Digital SAT Suite*

The previous subsection builds an argument that it is essential that students leave secondary school prepared to work with data, armed with statistical thinking skills and data acumen. Additionally, students need to understand concepts from the study of probability in order to understand the importance of randomness in statistics.

Two foundational topics that flow through the math curriculum, typically starting in grade six and continuing through high school, are developing an understanding of proportional reasoning and applying proportional relationships to solve single-step and multistep problems. Proportional reasoning is an important skill when solving percent-based problems, including discounts, tips, sales tax, interest, unit rates, and percent increase and decrease, and thus it is assessed, at appropriately challenging levels, throughout the digital SAT Suite, including on the SAT.

The Problem-Solving and Data Analysis content domain assesses knowledge and skills in solving problems using ratios, rates, proportional relationships, unit analysis, percentages, probability and conditional probability, one- and two-variable data, scatterplots, and models. Unlike topics covered in the Algebra and Advanced Math content domains, the topics addressed by the digital SAT Suite in Problem-Solving and Data Analysis are not aligned to those covered in a specific secondary-level course. State education systems include the topics covered in this domain in a variety of courses, starting with middle school/junior high school math and continuing through high school. The test items in the Problem-Solving and Data Analysis domain range in difficulty from relatively easy to relatively complex and challenging and test a wide range of reasoning skills.

## Geometry and Trigonometry

### *Why Geometry and Trigonometry Are Important*

Because geometry, historically the study of shapes and their properties, originates in the study of the measurement of the earth (Merriam-Webster, n.d.), it is one of the oldest branches of math and is, in some ways, the most immediately relevant. Freudenthal (1971) argued that the study of math should be tied to the world in which we live, else it is easily forgotten and rarely used. Geometry is inherently related to modeling the world around us, which includes measuring objects in space and developing spatial and deductive reasoning.

The value of geometry in K–12 education extends beyond the typical merits of understanding a subject to helping lay the foundations for achievement in other branches of math. Topics in geometry and measurement were considered Critical Foundations of Algebra by the National Mathematics Advisory Panel (2008). The report the panel produced specifically discussed the importance of similar triangles to understanding slope and linear functions. In addition, the panel suggested that to prepare for algebra, “students should be able to analyze the properties of two- and three-dimensional shapes using formulas to determine perimeter, area, volume, and surface area” and “should also be able to find unknown lengths, angles, and areas” (National Mathematics Advisory Panel, 2008, p. 18). Similarly, the authors of the Common Core State Standards for Mathematics (CCSSM) (NGA Center for Best Practices & CCSSO, 2010b) observed that “solving real-world and mathematical problems involving angle measure, area, surface area, and volume” are high priorities for college and career readiness (p. 84). Additionally, a survey of college math faculty (Er, 2018) rated reasoning and generalization—skills developable through the study of geometry—as both the most important math competencies for incoming college students to have previously mastered and the least likely to have been attained. Further, the study of geometry prepares students for trigonometry and precalculus, the latter of which Atuahene and Russell (2016) have shown that 53% of first-year college students struggle with, earning D, F, or W (withdrawal) grades at the end of a semester-long course. Geometry and trigonometry content is important not only academically for STEM fields (e.g., engineering, medicine) but also for careers in the trades (e.g., transportation, construction) and the arts (Morgan, 2018).

Historical assessment data from students in the United States relative to students from other nations show a long-term trend of weak performance on items related to geometric reasoning and measurement (Carpenter et al., 1980; Fey, 1984; Stigler et al., 1990). More recent findings have not improved the picture. Data from the Trends in International Mathematics and Science Study (TIMSS) highlighted geometry and measurement as the biggest areas of weakness for eighth-grade students from the U.S. (Ginsburg et al., 2005), and geometry performance by U.S. high school students was the lowest among the 16 participating countries (Mullis et al., 1998). The most recent National Assessment of Educational Progress (NAEP) with publicly released test items (National Assessment Governing Board, 2013) includes a grade 12 item asking students to determine the area of a triangle in a 3D figure. Only 5% of U.S. students were able to give a correct answer and show how they found the area of the figure. Additionally, it is well documented that U.S. high school students struggle with formal proof (e.g., Stylianides et al., 2017), which is why they need more opportunities for informal reasoning and sense making. These data are concerning given the importance of these topics for college and career readiness.

***Geometry and Trigonometry on the Digital SAT Suite***

Geometry is all about modeling the world around us, and knowledge of geometry helps lay the foundation for further achievement in math. Skills, knowledge, and concepts learned in the study of geometry are included in items in the Geometry and Trigonometry content domain (for the PSAT 8/9 only, the Geometry domain) but are also woven into items in the Algebra and Advanced Math domains, where geometric objects are sometimes used as contexts for building functions or modeling real-world scenarios. Geometry content on the digital SAT Suite is covered in secondary-level courses from grade 6 through high school. Trigonometry skills and knowledge are tested only on the SAT, PSAT/NMSQT, and PSAT 10, as these are typically taught and learned only in more advanced high school courses.

Test items in the Geometry and Trigonometry content domain involve applying skills and knowledge in finding areas, perimeters, volumes, and surface areas; using concepts and theorems related to lines, angles, and triangles (PSAT 8/9 includes triangle angle sum theorem only); solving problems using right triangles (SAT, PSAT/ NMSQT, and PSAT 10 only); solving problems using right triangle trigonometry (SAT, PSAT/NMSQT, and PSAT 10 only); calculating using sine, cosine, and tangent (SAT only); solving problems using radian measure and trigonometric ratios in the unit circle (SAT only); and using definitions, properties, and theorems relating to circles (SAT only). These test items vary in difficulty from easy to very hard and allow students to demonstrate problem-solving skills and knowledge using a variety of solving strategies.

## 8.3 Response Process–Oriented Validity Evidence

### 8.3.1 Cognitive Labs

In 2023, College Board undertook a cognitive lab study to confirm that, as with the paper and pencil SAT (College Board & HumRRO, 2020), items on the digital SAT Suite are capable of eliciting from students the sorts of higher-order, cognitively complex thinking required for college and career readiness. The methodology of this study made use of think-aloud protocols to gain insight into students' thinking processes as they read and answer select digital SAT Suite test items. Such evidence is important, first, because it would serve to confirm that the digital SAT Suite is an appropriately challenging set of assessments aligned with college and career readiness requirements and, second, because federal peer review of state accountability systems using the digital SAT Suite requires such evidence for the states' systems to meet expectations.

For the 2023 study (College Board, 2024), 26 high school juniors and seniors were instructed to think aloud as they answered a set of 20 Reading and Writing section items, while another 23 students participated in thinking aloud through a set of 20 Math section items. Items from both sections were chosen to be broadly representative of the sections' designs, including key skill/knowledge elements, item difficulty levels, subject areas, item formats (for Math), and text complexity levels (for Reading and Writing). Each participant engaged in a one-on-one interview session conducted via Zoom, wherein students were briefed on the task by a trained interviewer, experienced modeling of thinking aloud by the interviewer, had one or more opportunities to practice thinking aloud themselves, and then conveyed as much as possible about their concurrent thoughts as they worked through and attempted to answer a set of digital SAT Suite test items. Transcripts were produced from these interview sessions and analyzed qualitatively and quantitatively by College Board assessment and subject matter experts.

Qualitatively, each student's response to each test item was coded against a set of required (Reading and Writing) or expected (Math) behaviors. These behaviors, predefined by the College Board research team, described the aspects of cognitively complex thinking various item types are intended to elicit. Each student participant was judged by the researchers to have or have not demonstrated each of these behaviors in their response to the items, and their responses were coded correspondingly. Vignette candidates of students exhibiting these behaviors and, in the process, demonstrating exemplary (if not necessarily perfect) thinking through a given item were also identified during the coding stage.

In quantitative terms, College Board researchers tabulated several statistics from the coding. The most important metric for each Reading and Writing and Math item is referred to as the *differential*. This differential is the arithmetic difference between (1) the number of students who answered a given test item correctly and (2) the number of students who both answered the item correctly and also demonstrated all required (Reading and Writing) or at least one expected (Math) behavior. A low differential—one of 5 or lower—was deemed evidence of a given test item having performed as intended, as the majority of students would have demonstrated requisite elements of cognitively complex thinking in line with the item type's intended construct (i.e., the academic concept the item type is trying to assess students' attainment of). A higher differential, by contrast, was suggestive that a given item was not performing as intended, though mitigating factors may have led the researchers to conclude that the item was still capable of eliciting aspects of cognitively complex thinking.

All examined Reading and Writing items and the vast majority (85%) of examined Math items performed as intended, with differentials from 0 to 5. Two Math items had differentials greater than 5, but the qualitative evidence suggests that students were still exhibiting aspects of cognitively complex mathematical reasoning. A third Math item was answered correctly by no student, so although it technically had a differential of 0, it was considered an outlier. Vignettes of student performance associated with each of the 40 items supply additional evidence that the items elicited cognitively complex thinking from student participants.

The key finding of this study is strong confirmation of the hypothesis that the digital SAT Suite assessments are capable of eliciting cognitively complex thinking from student test takers. This is important because, first, a large body of evidence supports the conclusion that students need to be able to engage in such thinking to be college and career ready (i.e., prepared to succeed in college or workforce training programs without remediation) and, second, because the U.S. Department of Education requires states using the digital-suite tests (or other off-the-shelf large-scale standardized assessments) as part of their education accountability systems to supply evidence that the tests are capable of eliciting such thinking. Based on the findings reported here, policymakers should have high confidence that the tests of the digital SAT Suite of Assessments satisfy these criteria. In addition, the results and the methodology laid out in this report may be useful to researchers interested in evaluating the cognitive demands of large-scale standardized assessments.

The full report can be downloaded from College Board's website at **satsuite.collegeboard.org/media/pdf/digital-sat-cognitive-lab-report.pdf**.

Beginning in 2024, College Board will conduct additional cognitive labs with members of select test-taking population subgroups, including English learners, students with specific learning disorder: reading (dyslexia), and students with ADHD, to examine these students' thought processes as they take portions of the tests. The two main goals here are, first, to learn more about how students in these population subgroups engage with test materials and, second, to see whether changes in test design introduced by the digital SAT Suite contribute to more accurate assessment of these students' knowledge and skills via the reduction or elimination of construct-irrelevant barriers. Findings from these studies will be shared in 2025 and will feed into future test design and development.

## 8.3.2 Usability/Accessibility Research

The following subsections discuss two studies conducted by College Board to evaluate and improve on the more complex of the two selected-response formats used for digital SAT Suite test items as well as the Bluebook application itself.

### Student-Produced Response (SPR) Item Format (Math Section Only)

College Board has extensively examined how best to implement the student-produced response (SPR) item format in the digital SAT Suite's Math section. This matter is of particular concern as the format requires more complex answer entry and verification steps on students' part than are required with selecting multiple-choice answer options.

The SPR format, which is used for roughly 25% of Math section items on any given digital-suite test form, is intended to complement the four-option, single-select multiple-choice format used for the remainder of the items (as well as for all Reading and Writing items). While both item formats are suitable for assessing a wide range of skills and knowledge in math, SPR items differ from multiple-choice items in that students must derive and enter their own answers rather than select from a predefined option set. The use of the SPR format thus serves to assess whether students can apply their math skills and knowledge to a variety of math problems without the scaffolding and support of multiple-choice answers.

Math SPR items require students to enter answers of up to six characters, the first of which may be a negative sign. For answers exceeding this limit, students are instructed to either round or truncate their results and are given examples of how to perform each. Students are also advised on how to properly enter fractional and decimal answers (as well as integer answers).

Over a lengthy period of development, feedback collection, and iterative improvement, College Board solicited input from content, measurement, and user experience stakeholders and students on a range of issues related to the SPR format. Across four phases of study (involving samples of 451, 796, 847, and 1,166 high school juniors and seniors), College Board researchers tested a range of SPR formats and features, including whether a single entry field or separate fields for each character should be supplied; whether students should be asked to fill in boxes or blanks; whether students preferred and had better success with dropdowns (for individual-character fields), onscreen keypads, and/or keyboards; whether the directions should be open by default or closed; how the directions themselves should appear; and what sorts of validation and error messaging would be most beneficial.

These studies concluded that having a single entry field for answers (rather than separate fields for each character) and the directions open by default provided the best results and experience. Bluebook, the digital testing application, also previews entered answers for students to help ensure that what they actually entered was what they had intended to enter (which is particularly important for the entry of mixed numbers), and error messages are presented when students make clear entry errors (e.g., a negative sign in a position other than as the first character).

### Bluebook Usability and Accessibility for Students Who are Blind or Visually Impaired

The Services for Students with Disabilities (SSD) department and the Accessibility Compliance Office (ACO), with support from other College Board members, conducted usability and accessibility testing at the 2022 National Federation of the Blind (NFB) convention on July 6–7, 2022. The NFB is the oldest and largest advocacy organization in the United States for blind and visually impaired people. The purpose of the test was to assess the usability and accessibility of the Bluebook application and sample test content. Because this study was conducted in 2022, this sample content consisted of the AP World History preview test, which uses Bluebook and includes complex graphics, such as maps.

The usability and accessibility testing was administered to a sample of blind and low-vision high school and post–high school attendees at the convention as well as to registered remote attendees participating in a virtual experience. Eleven conference attendees participated on day 1 (July 6), and 23 participated on day 2 (July 7). In addition, a total of nine virtual conference attendees participated across the two days.

According to the survey responses, 75% of NFB participants found the testing experience to be excellent/good, with 79% indicating that they were comfortable testing on their device. Respondents rated the alternate text descriptions highly for appropriateness (76%) and for providing enough information (86%).

Among the common observations were the following:

- Most participants liked the question navigator and the “mark for review” feature, finding these tools to be highly usable, intuitive, easy to navigate, and readable.
- Most users wanted the keyboard shortcuts for navigating the exam, entering responses, and using the exam features to be more discernable.
- Participants reported being overwhelmed with verbose alt text descriptions and indicated a desire for tactile graphics for complex images.
- Participants noted that test directions should be relevant for screen reader users. For example, the directions included information about using scratch paper and the countdown timer turning red when five minutes of testing remain, neither of which are of any use to a blind student. (Screen reader users can instead use the annotate tool to take notes during testing and set a notification for the five-minute warning.)
- Low-vision users found that test graphics did not scale while zooming.

College Board continues to make iterative improvements to Bluebook, to better address the needs and preferences of blind and low-vision test takers as well as users in general. For instance, since this study was conducted, College Board staff have devoted considerable attention to codifying and refining alt text style to make these verbal descriptions of visual images more concise, more precise, and easier to use. (It should also be noted that the digital SAT Suite tests do not use graphics as complex as those from the sample AP content tested in the 2022 NFB study.) College Board develops its alt text descriptions in partnership with content experts and in accordance with various professional standards, including the DIAGRAM Center’s Image Description Guidelines (**diagramcenter.org/table-of-contents-2.html**), NWEA’s Image Description Guidelines for Assessments (**nwea.org/accommodations-accessibility**), and the Web Content Accessibility Guidelines (WCAG) produced by WC3 Web Accessibility Initiative (**w3.org/WAI/standards-guidelines/wcag**). College Board also now strongly recommends that blind and low-vision students using screen readers request supplemental raised line drawings.

## 8.4 Relationship Between the SAT and Other Educational Measures

As mentioned earlier in this chapter, according to the *Standards for Educational and Psychological Testing*, one of the five primary sources of test score validity evidence is that based on relations to other variables (AERA et al., 2014). While most frequently those other variables are relevant outcomes or criteria of interest, there is also value in collecting what is referred to as *convergent evidence*, or relationships between test scores and other measures that assess similar constructs (e.g., academic preparation, college readiness). This section describes a study that examines relationships between digital SAT scores and other relevant educational measures, such as high school grade point average (HSGPA), PSAT/NMSQT total score, and average AP Exam score, and compares those relationships to paper and pencil SAT score relationships with the same measures.

This study helps clarify the nature and meaning of digital SAT scores, how they relate to paper and pencil SAT scores, and how digital SAT scores relate to various other educational measures. For the full version of these findings, please refer to *Digital SAT Score Relationships with Other Educational Measures* (Marini et al., 2022), from which this section has been adapted.

### 8.4.1 Sample

The sample of students for this study comes from a pilot concordance study conducted in spring 2022 in which high school students were invited to take both a digital SAT and paper and pencil SAT. This initial sample included 6,373 students.[10] Three subsamples were also created: one for those with a self-reported high school grade point average (n=6,160), one for those with PSAT/NMSQT total scores (n=5,638), and one for those with AP Exam scores (n=5,171). Table 8.4 provides more information about the characteristics of the full sample of students. The study sample tended to be female, primarily White, Asian, or Hispanic; have parents with a bachelor's degree or higher; and state that English is their best language. While this sample is not representative of the general digital SAT test-taking population, these demographic characteristics closely resemble College Board's validity research samples in larger studies of college-going students.

10 Students had to be in their junior year and had to have taken the paper and pencil SAT in March, April, or May 2022. As student motivation was a concern, students with more than a 200-point difference between their paper and pencil and digital SAT Reading and Writing and/or Math section scores (n=18) were removed. The researchers also removed one student with a Math section score of 200 on the paper and pencil SAT and a corresponding PSAT/NMSQT Math section score that was more than 200 points higher.

**TABLE 8.4** STUDENT DEMOGRAPHICS OF THE CONVERGENT VALIDITY STUDY SAMPLE

| Student Characteristic | Overall Sample (n=6,373) | |
|---|---|---|
| **Gender** | Male | 40% |
| | Female | 59% |
| | Another/No Response | <1% |
| **Race/Ethnicity** | American Indian or Alaska Native | <1% |
| | Asian | 24% |
| | Black or African American | 7% |
| | Hispanic or Latino | 21% |
| | Native Hawaiian or Other Pacific Islander | <1% |
| | White | 28% |
| | Two or More Races | 4% |
| | Not Stated | 16% |
| **Highest Parental Education Level** | No High School Diploma | 4% |
| | High School Diploma | 13% |
| | Associate Degree | 4% |
| | Bachelor's Degree | 36% |
| | Graduate Degree | 38% |
| | Not Stated | 5% |
| **Best Language** | English Only | 68% |
| | English and Another Language | 29% |
| | Another Language | 3% |
| | Not Stated | 1% |

*Note.* Percentages may not sum to 100 due to rounding.

## 8.4.2 Measures

**Paper and Pencil SAT Scores.** Paper and pencil SAT scores were obtained from College Board's database and matched to each student who participated in the special administration of the digital SAT. The paper and pencil SAT scores included in this study were the following:

*SAT total score (400–1600 scale, in 10-point increments)*: Sample mean of 1242 (standard deviation [SD]=166)

*SAT Evidence-Based Reading and Writing (ERW) section score (200–800 scale, in 10-point increments)*: Sample mean of 621 (SD=81)

*SAT Math section score (200–800 scale, in 10-point increments)*: Sample mean of 621 (SD=98)

**Digital SAT Scores.** A special administration of the digital SAT was given as part of the first concordance study. Digital SAT scores were obtained from College Board records and matched to every student who participated in the study. The digital SAT scores included in this study were the following:

*SAT total score (400–1600 scale, in 10-point increments)*:
Sample mean of 1241 (SD=145)
*SAT Reading and Writing section score (200–800 scale, in 10-point increments)*:
Sample mean of 621 (SD=71)
*SAT Math section score (200–800 scale, in 10-point increments)*:
Sample mean of 620 (SD=88)

**High School GPA (HSGPA).** Self-reported HSGPA was obtained from the SAT Questionnaire completed when students registered for the SAT and is reported on a 12-point scale ranging from 0.00 (F) to 4.33 (A+). The HSGPA measure in this study has a sample mean of 3.90 (SD=0.41).

**PSAT/NMSQT Total Score (320–1520 scale, in 10-point increments).** PSAT/NMSQT total scores were obtained from College Board's database and matched to each student who participated in the special administration of the digital SAT. The PSAT/NMSQT total score in this study had a sample mean of 1178 (SD=159).

**Average AP Exam Score.** AP Exam scores were obtained for each student in the sample, and average AP Exam score was calculated from all AP Exams a given student took. This average ranged from 1 to 5, with a sample mean of 3.2 (SD=1.0).

Table 8.5 includes descriptive statistics for all measures of interest in this study. The descriptive statistics for HSGPA, PSAT/NMSQT total score, and average AP Exam score reflect the values for the measure-specific sample and not the full SAT sample.

**TABLE 8.5 DESCRIPTIVE STATISTICS FOR MEASURES OF INTEREST, OVERALL CONVERGENT VALIDITY STUDY SAMPLE**

| Study Measure | n | Mean | SD | Min. | Max. |
|---|---|---|---|---|---|
| Paper and Pencil SAT Total Score | 6,373 | 1242 | 166 | 600 | 1600 |
| Paper and Pencil SAT ERW Section Score | 6,373 | 621 | 81 | 310 | 800 |
| Paper and Pencil SAT Math Section Score | 6,373 | 621 | 98 | 260 | 800 |
| Digital SAT Total Score | 6,373 | 1241 | 145 | 670 | 1600 |
| Digital SAT Reading and Writing Section Score | 6,373 | 621 | 71 | 310 | 800 |
| Digital SAT Math Section Score | 6,373 | 620 | 88 | 270 | 800 |
| HSGPA | 6,160 | 3.90 | 0.41 | 1.67 | 4.33 |
| PSAT/NMSQT Total Score | 5,638 | 1178 | 159 | 510 | 1520 |
| Average AP Exam Score | 5,171 | 3.2 | 1.0 | 1 | 5 |

## 8.4.3 Methodology

College Board conducted correlational analyses using both paper and pencil SAT and digital SAT scores. First, the researchers examined the strength of the relationships between students' section and total scores on the paper and pencil SAT and the corresponding section and total scores on the digital SAT. Next, they examined the strength of the relationships between students' paper and pencil SAT and digital SAT scores with three other measures of interest: HSGPA, PSAT/NMSQT total score, and average AP Exam score. Although all students (n=6,373) had both paper and pencil SAT and digital SAT scores, not all students had HSGPAs or PSAT/NMSQT total and/or AP Exam scores. To maximize sample sizes, the researchers conducted separate analyses for each of the three measures of interest: HSGPA (n=6,160), PSAT/NMSQT total score (n=5,638), and average AP Exam score (n=5,171).

## 8.4.4 Results

### Paper and Pencil SAT and Digital SAT Relationships

Observed correlations between digital and paper and pencil SAT scores are presented in Table 8.6.[11] As expected, the strongest correlation for the digital SAT total score was with the paper and pencil SAT total score at .92. The digital SAT Reading and Writing section score most strongly correlated with the paper and pencil SAT ERW section score at .86; similarly, the digital SAT Math section score most strongly correlated with the paper and pencil SAT Math section score at .90. These correlations provide strong evidence of the convergent validity of digital SAT scores. Moreover, these correlations are consistent with test-retest correlations found between paper and pencil SAT scores for examinees who tested more than one time.[12] This result indicates that students are scoring as similarly on the digital SAT and paper and pencil SAT as they were when they sat for two different administrations of the paper and pencil SAT.

**TABLE 8.6 CORRELATIONS AND 95% CONFIDENCE INTERVALS BETWEEN DIGITAL SAT AND PAPER AND PENCIL SAT SCORES (N=6,373)**

| Paper and Pencil SAT Score | Digital SAT Score | Correlation (95% Confidence Intervals) |
|---|---|---|
| Total | Total | .92 (.92–.93) |
| ERW Section | Reading and Writing Section | .86 (.85–.86) |
| Math Section | Math Section | .90 (.89–.90) |

*Paper and Pencil and Digital SAT Relationships with Related Measures*

11 Mean values for the academic variables studied are higher for this study sample than for the general SAT test-taking population, with smaller standard deviations. The smaller standard deviations denote a restriction of range in this sample. Therefore, a sample that more closely resembled the general population would likely have correlations that are slightly higher than those found in this study.

12 For a sample (n=41,120) of students who took the paper and pencil SAT in both March and May 2022, the test-retest reliability estimates for SAT total, ERW section, and Math section scores were .93, .88, and .90, respectively.

Figure 8.1 illustrates the strength of the relationships between digital and paper and pencil SAT scores and the three educational measures of interest: HSGPA, PSAT/NMSQT total score, and average AP Exam score. Both correlations and 95% confidence intervals are provided. The differences in the relationships between both digital and paper and pencil SAT total scores and the three other measures were relatively minor, varying from .00 or .01. The same was true for SAT Math section scores. Differences in correlations between paper and pencil SAT ERW and digital SAT Reading and Writing section scores with the other three measures ranged from .01 (average AP Exam scores) to .04 (PSAT/NMSQT total score).

**FIGURE 8.1 DIGITAL SAT AND PAPER AND PENCIL SAT CORRELATIONS WITH HSGPA, PSAT/NMSQT TOTAL SCORE, AND AVERAGE AP EXAM SCORE**

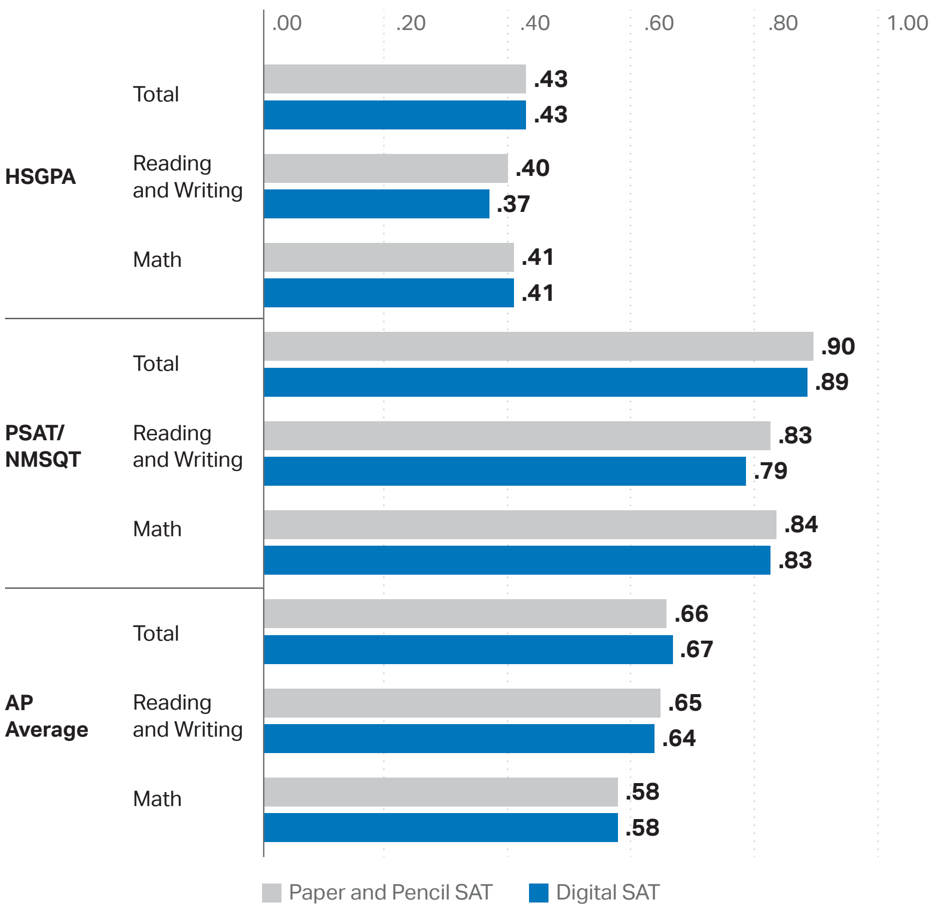


Table 8.7 presents correlations between digital and paper and pencil SAT scores with the three educational measures of interest along with 95% confidence intervals. With one exception, the confidence intervals overlap.[13] This suggests that going forward, the relationships between the digital SAT and these three measures of educational achievement will be as strong as those found with the paper and pencil SAT.

**TABLE 8.7 DIGITAL SAT AND PAPER AND PENCIL SAT CORRELATIONS WITH HSGPA, PSAT/NMSQT TOTAL SCORE, AND AVERAGE AP EXAM SCORE**

| Educational Measure | Test Score | n | Paper and Pencil SAT Correlation (95% Confidence Interval) | Digital SAT Correlation (95% Confidence Interval) |
|---|---|---|---|---|
| HSGPA | Total | 6,160 | .43 (.41, .45) | .43 (.41, .45) |
| | (Evidence-Based) Reading and Writing Section | 6,160 | .40 (.38, .42) | .37 (.35, .39) |
| | Math Section | 6,160 | .41 (.39, .43) | .41 (.39, .43) |
| PSAT/NMSQT Total Score | Total | 5,638 | .90 (.89, .90) | .89 (.89, .90) |
| | (Evidence-Based) Reading and Writing Section | 5,638 | .83 (.82, .84) | .79 (.78, .80) |
| | Math Section | 5,638 | .84 (.83, .85) | .83 (.82, .84) |
| Average AP Exam Score | Total | 5,171 | .66 (.65, .68) | .67 (.65, .68) |
| | (Evidence-Based) Reading and Writing Section | 5,171 | .65 (.63, .66) | .64 (.62, .65) |
| | Math Section | 5,171 | .58 (.56, .60) | .58 (.56, .59) |

## 8.4.5 Summary

Results from this study indicate that students' digital SAT scores are strongly related to their scores on the paper and pencil SAT. Moreover, the strength of the relationships of the digital SAT with other measures of educational achievement—HSGPA, PSAT/NMSQT total score, and average AP Exam score—parallel the strength of the relationships found between the paper and pencil SAT and these measures.

13 The scores used in this study were from the paper and pencil PSAT/NMSQT and not the digital PSAT/NMSQT. We would expect the digital SAT section and total scores to be more strongly correlated with the corresponding digital PSAT/NMSQT section and total scores as they are assessed in the same modality.

# 8.5 Relationship Between the SAT and College Outcomes

In accordance with the *Standards for Educational and Psychological Testing* (AERA et al., 2014), modifications to an exam necessitate an analysis of how the new scores relate to the outcomes they are intended to predict. For the digital SAT, this includes the prediction of college academic performance to inform the understanding of the utility of digital SAT scores in college admission, placement, scholarship, and advising decisions and processes on campus. The study reported on in this section examines predictive relationships between digital SAT scores and first-year college GPA (FYGPA). Note that another similar but smaller-scale study was conducted on digital SAT score relationships with first-semester college outcomes (Marini et al., 2023). For the full version of these findings please refer to *Digital SAT Pilot Predictive Validity Study: A Comprehensive Analysis of First-Year College Outcomes* (Westrick et al., 2023), from which this section has been adapted.

## 8.5.1 Methodology

### Study Design

The aim of this study was to recruit 10 to 15 diverse four-year institutions so that students (75 to 250 per campus) could then be recruited to participate in an administration of the digital SAT very early in their first year of college. College Board offered students $150 gift cards for participating in the exam and an additional $50 incentive if their digital scores met or exceeded their PSAT/NMSQT or SAT scores on record at College Board. Student participants also agreed to have their institutions share their first-year college performance information with College Board.

### Institutional Sample

The desired sample of institutional participants was intended to reflect the population of four-year higher education institutions as closely as possible while also facilitating a successful study (e.g., focusing on larger institutions that would be more likely to recruit substantial numbers of student participants). Ultimately, 12 four-year institutions were recruited for the initial study. Institutional participants were more likely to be public and very large institutions. The institutions in the study varied by admittance rate and U.S. geographic area and included one historically Black college/university (HBCU) and two Hispanic-serving institutions (HSIs). One of the 12 institutions did not provide sufficient coursework data and was therefore excluded from the study sample. The characteristics of the 11 institutions ultimately included in the analyses are summarized in Table 8.8.[14]

14 See Appendix Table B-1 for more information on the institutional population.

**TABLE 8.8 INSTITUTIONAL CHARACTERISTICS OF THE PILOT PREDICTIVE VALIDITY STUDY SAMPLE**

| Institutional Characteristic | Percent of Total Sample (k=11) | |
|---|---|---|
| U.S. Region | Midwest | 9% |
| | Mid-Atlantic | 9% |
| | New England | 18% |
| | South | 27% |
| | Southwest | 18% |
| | West | 18% |
| Control | Public | 64% |
| | Private | 36% |
| Admittance Rate | Under 25% | 27% |
| | 25% to 50% | 9% |
| | 51% to 75% | 55% |
| | Over 75% | 9% |
| Undergraduate Enrollment | Small (n<5,000) | 0% |
| | Medium (5,000<n<9,999) | 0% |
| | Large (10,000<n<19,999) | 9% |
| | Very Large (n≥20,000) | 91% |

*Note*. Percentages by institutional characteristic may not sum to 100 due to rounding.

## Student Sample

A total of 1,990 first-year, first-time college students participated in the digital SAT pilot exam administrations across the original 12 institutions in the study.[15] All students had graduated from high school in spring 2022 and had prior SAT or PSAT/NMSQT scores on record at College Board. The study inclusion criteria required that students have a self-reported HSGPA and a FYGPA, and students who experienced any section score decrease of 200 points or more from paper and pencil to digital SAT testing, thereby indicating questionable motivation, were excluded, resulting in a final sample of 1,889 students. Demographic information regarding the study sample is presented in Table 8.9. The sample included more female than male students and about one-third underrepresented minority students, one-third Asian students, and one-third White students. Slightly more than one-fourth of the sample reported that English and another language or a language other than English was their best language, and most students had parents with a bachelor's or graduate degree.[16] See Appendix Table B-2 for demographic information for the members of

15 One institution did not provide study participants' first-year grades to the College Board and therefore could not be included in this study.

16 As a check, College Board researchers reweighted the sample to more closely resemble typical SAT validity study populations in terms of institutional and student characteristics. They found that these new correlations were all within the 95% confidence intervals of the original sample correlations; therefore, analyses were conducted on the original sample.

the high school class of 2022 who took the SAT and the most recent national SAT validity study sample that, like this study sample, includes only enrolled college students (from the entering class of fall 2020).

**TABLE 8.9 STUDENT CHARACTERISTICS OF THE PILOT PREDICTIVE VALIDITY STUDY SAMPLE**

| Student Characteristic | Percent of Total Sample (n=1,889) | |
|---|---|---|
| Gender | Male | 42% |
| | Female | 58% |
| | Another/Omitted | <1% |
| Ethnicity | American Indian/Alaska Native | <1% |
| | Asian | 33% |
| | Black/African American | 6% |
| | Hispanic/Latino | 21% |
| | Native Hawaiian/Other Pacific Islander | <1% |
| | White | 33% |
| | Two or More Races | 4% |
| | No Response | 3% |
| Best Language | English Only | 74% |
| | English and Another Language | 24% |
| | Language Other Than English | 2% |
| | No Response | <1% |
| Highest Parental Education Level | No High School | 4% |
| | High School Diploma | 14% |
| | Associate Degree | 4% |
| | Bachelor's Degree | 35% |
| | Graduate Degree | 39% |
| | No Response | 3% |

*Note.* Percentages by student characteristic may not sum to 100 due to rounding.

## 8.5.2 Measures

**Paper and Pencil SAT Scores.** Official paper and pencil SAT scores were obtained from College Board's database and matched to each student who participated in the special administrations of the digital SAT. The paper and pencil SAT scores included in this study were the following:

*SAT total score (400–1600 scale, in 10-point increments)*:
Sample mean of 1332 (SD=158)

*SAT Evidence-Based Reading and Writing (ERW) section score (200–800 scale, in 10-point increments)*: Sample mean of 662 (SD=76)

*SAT Math section score (200–800 scale, in 10-point increments)*:
Sample mean of 669 (SD=94)

**Digital SAT Scores.** Special administrations of the digital SAT took place at the 12 participating college campuses over 4 weekends in September and October 2022. The digital SAT scores included in this study were the following:

*SAT total score (400–1600 scale, in 10-point increments)*:
Sample mean of 1297 (SD=163)

*SAT Reading and Writing section score (200–800 scale, in 10-point increments)*:
Sample mean of 643 (SD=85)

*SAT Math section score (200–800 scale, in 10-point increments)*:
Sample mean of 654 (SD=95)

**High School GPA (HSGPA).** Self-reported HSGPA was obtained from the SAT Questionnaire completed when students registered for the SAT (or PSAT/NMSQT) and is reported on a 12-point scale ranging from 0.00 (F) to 4.33 (A+). The HSGPA measure in this study had a sample mean of 3.97 (SD=0.34).

**First-Year Credits Earned (FYCE).** Course credits completed in all courses in the first year of college were obtained from the participating institutions. If a student failed a course, the credits earned equaled zero. The sample mean was 29 (SD=6). Note that most bachelor's degree programs require 120 college credits to graduate.

**First-Year College GPA (FYGPA).** First-year GPA and grades in all courses in the first year of college were obtained from the participating institutions. FYGPA was reported on a 0.00 to 4.00, continuous scale. The sample mean FYGPA was 3.59 (SD=0.49).

**Domain-Specific GPAs.** All college courses were coded for content area so that analyses could be conducted on domain-specific grade point averages. The three domain-specific college GPAs in the current study were math GPA (n=1,384, mean=3.34, SD=0.85), STEM GPA (science, technology, engineering, and math; n=1,765, mean=3.45, SD=0.68), and all-but-math GPA (n=1,889, mean=3.61, SD=0.50), the last serving as a criterion for analyses with the Reading and Writing section (as most courses in college involve reading and writing).

Domain-specific grade point averages were calculated within student and across all relevant course grades received in a particular area during the first year of college, excluding remedial coursework. For example, if a student took only one math course in their first year, their average course grade in math would be based on the grade earned in that one course. If a student took three math courses, the average course grade would be based on the average of the three course grades earned, with these calculations taking into account both the grades earned in each course and the number of credits associated with each course.

## Descriptive Statistics

Table 8.10 shows the descriptive statistics for the measures included in the study. First-year GPA for students who did not complete the second semester were based on the courses they did complete in the first semester. As is typical in predictive validity research involving enrolled college students (e.g., Shaw et al., 2016; Westrick et al., 2019), the sample in this study was academically quite strong.

**TABLE 8.10 DESCRIPTIVE STATISTICS FOR THE PILOT PREDICTIVE VALIDITY STUDY**

| Educational Measure | n | Mean | SD | Min. | Max. |
|---|---|---|---|---|---|
| HSGPA | 1,889 | 3.97 | 0.34 | 1.67 | 4.33 |
| First-Year Credits Earned | 1,889 | 29 | 6 | 1 | 43 |
| First-Year GPA | 1,889 | 3.59 | 0.49 | 0.58 | 4.00 |
| First-Year Other-than-Math GPA | 1,889 | 3.61 | 0.50 | 0.38 | 4.00 |
| First-Year STEM GPA | 1,765 | 3.45 | 0.68 | 0.00 | 4.00 |
| First-Year Math GPA | 1,384 | 3.34 | 0.85 | 0.00 | 4.00 |
| Paper and Pencil SAT ERW Section Score | 1,889 | 662 | 76 | 400 | 800 |
| Digital SAT Reading and Writing Section Score | 1,889 | 643 | 85 | 330 | 800 |
| Paper and Pencil SAT Math Section Score | 1,889 | 669 | 94 | 360 | 800 |
| Digital SAT Math Section Score | 1,889 | 654 | 95 | 330 | 800 |
| Paper and Pencil SAT Total Score | 1,889 | 1332 | 158 | 830 | 1600 |
| Digital SAT Total Score | 1,889 | 1297 | 163 | 680 | 1600 |

## 8.5.3 Analysis

Study analyses included correlational analysis to arrive at the incremental utility (Schmidt & Hunter, 1998) gained when SAT scores are used alongside HSGPA to predict first-year college grades and course credits. Analyses were conducted at the institution level, and then the results were weighted by institution size (number of students), aggregated, and then averaged using the total number of students. As admission selectivity restricts the range of students enrolled at the institutions, College Board researchers followed standard practices to statistically correct the raw correlations because these typically underestimate the true relationship between test scores and college outcomes (AREA et al., 2014).[17] They also followed the standard practice of reporting both raw and adjusted correlations. In addition to correlations, graphical depictions of mean differences in FYGPA, domain-specific GPA, or course credits by SAT total score bands were used. For dichotomized outcomes—earning a FYGPA of 3.0 or higher and earning 30 or more credits in the first year—logistic regression analyses were conducted at the institution level before weighting and aggregating to arrive at average parameter estimates.

## 8.5.4 Results

Table 8.11 displays the intercorrelations between digital SAT Reading and Writing section scores, digital SAT Math section scores, and HSGPA. Consistent with previous research (Kobrin et al., 2008; Shaw et al., 2016), the correlations between each of the digital SAT section scores and HSGPA are approximately .50, indicating a strong relationship but also that the digital SAT and HSGPA are not identical constructs and therefore offer unique and complementary information about a student.

**TABLE 8.11 CORRECTED (RAW) CORRELATION MATRIX OF DIGITAL SAT SCORES AND HSGPA IN THE PILOT PREDICTIVE VALIDITY STUDY**

| Educational Measure | SAT Reading and Writing Section Score | SAT Math Section Score |
|---|---|---|
| SAT Reading and Writing Section Score | | |
| SAT Math Section Score | .81 (.63) | |
| HSGPA | .49 (.25) | .50 (.26) |

*Note.* The correlation between digital SAT total score and HSGPA was .52 (.28).

17 Without information on how students who were not admitted or who did not enroll would have performed at an institution, only a partial glimpse into how the tests work for selection is possible. This circumstance restricts the variability (range) in test scores available for analysis since the available test scores tend to be the higher scores of those students who were admitted and did enroll, minimizing the test score–criterion relationship. Correlations in this study were corrected for multivariate range restriction (Lawley, 1943) using the 2022 graduating seniors who took the SAT as the reference population.

## First-Year GPA

Table 8.12 presents the correlations between predictors (digital SAT scores and HSGPA) and FYGPA. Individually, the SAT and HSGPA had strong relationships with FYGPA, with correlations of .57 and .54, respectively, and jointly they had an even stronger relationship as indicated by a multiple correlation of .66, a 22% increase in predictive utility over using HSGPA alone.[18] For perspective, correlations with absolute values of .50 or higher are considered large (Cohen, 1988), indicating a strong relationship between SAT scores and FYGPA.

**TABLE 8.12 CORRECTED (RAW) CORRELATION OF PREDICTORS WITH OVERALL FIRST-YEAR GPA (K=11, N=1,889)**

| Predictor(s) | Correlation | 95% Confidence Interval |
|---|---|---|
| Digital SAT Reading and Writing Section Score | .53 (.32) | .50–.56 |
| Digital SAT Math Section Score | .55 (.35) | .52–.58 |
| Digital SAT Total Score | .57 (.39) | .54–.60 |
| HSGPA | .54 (.27) | .51–.57 |
| Digital SAT+HSGPA | .66 (.46) | .63–.68 |
| Digital SAT incremental validity beyond the use of HSGPA alone | .12 (.19) | |

*Note.* Confidence intervals were calculated using the adjusted correlations after rounding.

The positive relationships between digital SAT scores and FYGPA presented in Table 8.12 can be better understood when presented visually. Figure 8.2 demonstrates the relationship between digital SAT total scores and FYGPA. This figure depicts a clear, strong, and positive relationship between digital SAT scores and FYGPA: as it moves to higher digital SAT score bands, mean FYGPA increases. For example, students earning digital SAT total scores of less than 1000 have a mean FYGPA of 3.19 in this study, while students earning digital SAT scores from 1500 to 1600 had, on average, a FYGPA of 3.84.

18 This value was calculated by subtracting the HSGPA-FYGPA correlation (.54) from the multiple correlation of HSGPA and SAT with FYGPA (.66) to arrive at the SAT incremental validity coefficient (.12). This coefficient is then divided by the HSGPA-FYGPA correlation (.54) and multiplied by 100 to arrive at the incremental predictive utility value of 22%. As there were differences between the sample and the typical national cohort, College Board researchers performed checks by reweighting the sample to more closely resemble typical validity study populations in terms of student characteristics. They found that reweighted correlations were all within the 95% confidence intervals of the sample correlations presented in Table 8.12.

**FIGURE 8.2** MEAN FIRST-YEAR GPA BY DIGITAL SAT TOTAL SCORE BAND

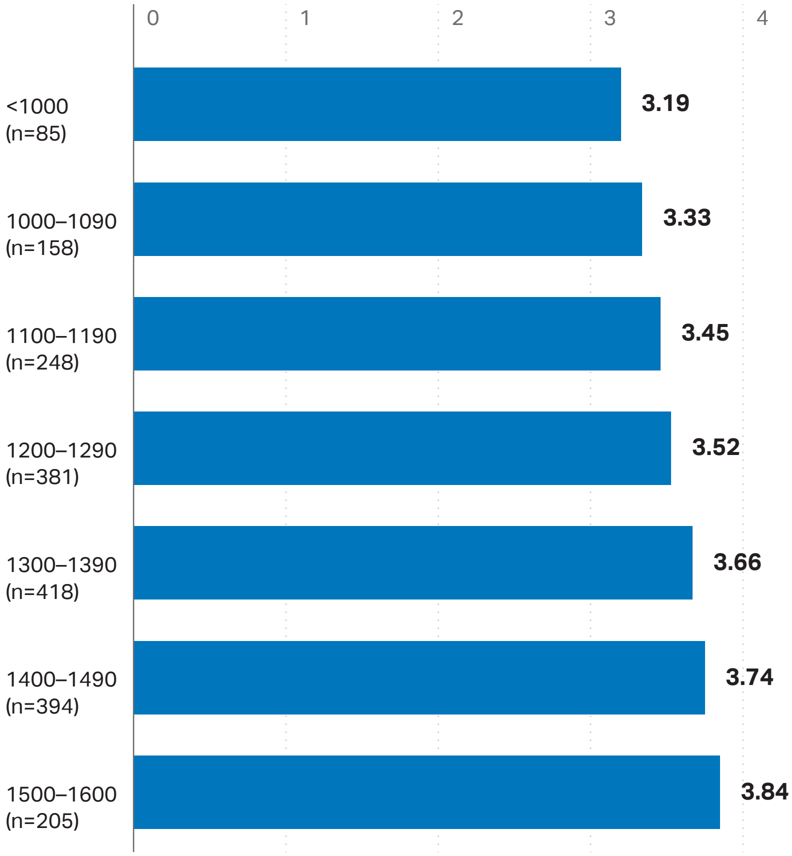
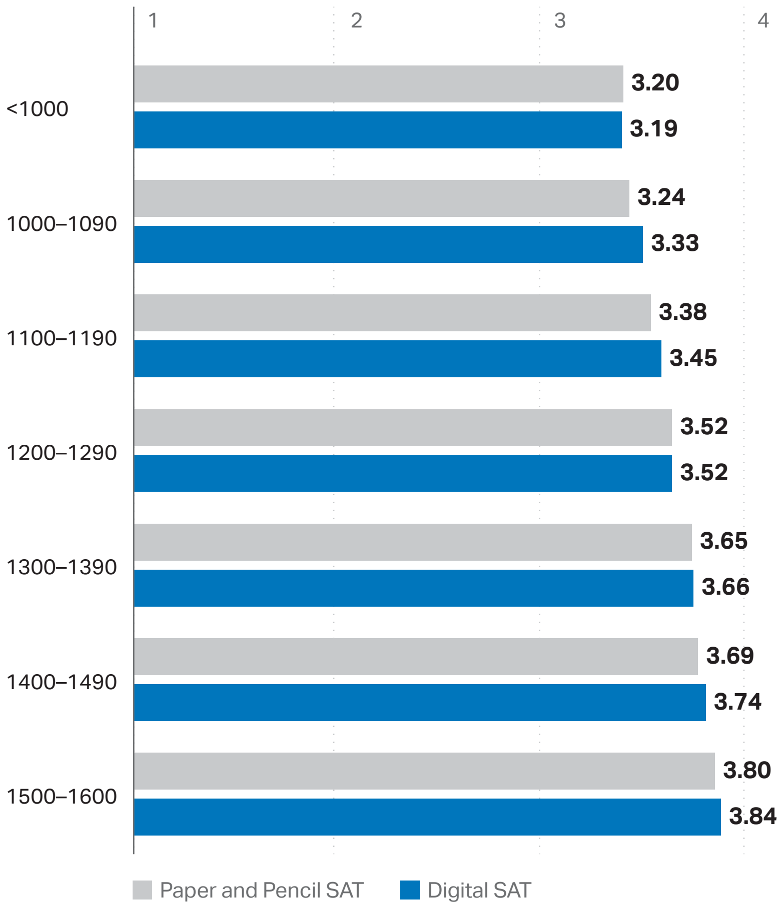


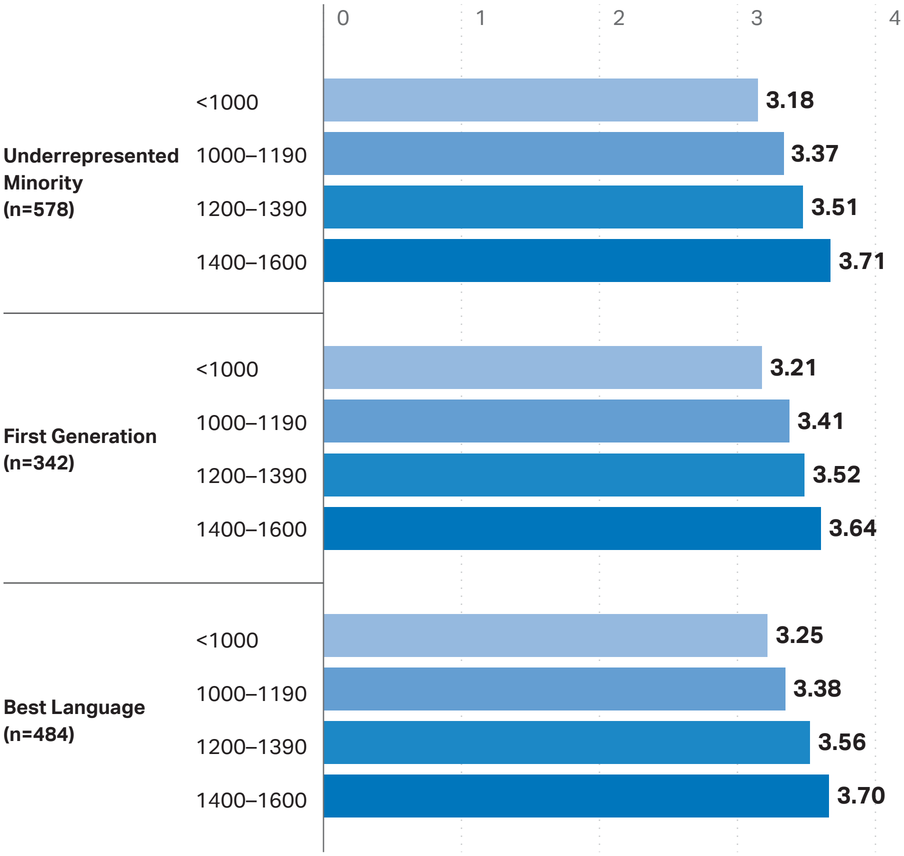

Figure 8.3 presents mean FYGPA across both the paper and pencil and digital SAT total score bands. Students' mean FYGPA increased in tandem with their SAT scores on both the paper and pencil and the digital SAT. Moreover, the relationship between digital SAT scores and FYGPA is nearly identical to that of paper and pencil SAT scores and FYGPA for these student. These results should assure SAT score users of the value of digital SAT scores for understanding student readiness for college and career; informing admission, course placement, and scholarship decisions; and identifying students needing academic support.

**FIGURE 8.3** MEAN FIRST-YEAR GPA BY PAPER AND PENCIL SAT AND DIGITAL SAT TOTAL SCORE BAND




To provide further validity evidence, College Board researchers conducted subgroup analyses. Results for three student subgroups of interest—underrepresented minority students, first-generation college students, and students whose self-identified best language(s) were either English and another language or a language other than English—can be found in Figure 8.4. As was observed for the overall sample, students' mean FYGPA increases as SAT score bands increase. The results presented in Figure 8.4 provide evidence of the value of digital SAT scores in understanding the college performance of these student subgroups.

**FIGURE 8.4** MEAN FIRST-YEAR GPA BY DIGITAL SAT TOTAL SCORE BAND: SUBGROUP ANALYSES




Controlling for HSGPA

Figure 8.5 illustrates how digital SAT total scores differentiate academic performance among groups of students with the same HSGPA, in essence controlling for HSGPA to indicate the added informational value of SAT scores. For example, for students with an "A" HSGPA, those with digital SAT total scores from 1000 to 1090 had a mean FYGPA of 3.33, while those with digital SAT total scores from 1400 to 1490 had a mean FYGPA of 3.76. Similar patterns are observable for students with HSGPAs of A– and A+. (College Board researchers focused this analysis on the more than 91% of the sample that reported a HSGPA of 3.67 or higher.) Figure 8.5 also represents the 22% improvement in predictive utility from using digital SAT scores in addition to HSGPA to predict students' first-year college performance, based on the correlations with FYGPA presented in Table 8.12.

**FIGURE 8.5** MEAN FIRST-YEAR GPA BY DIGITAL SAT TOTAL SCORE BAND WITHIN HSGPA FOR "A" STUDENTS

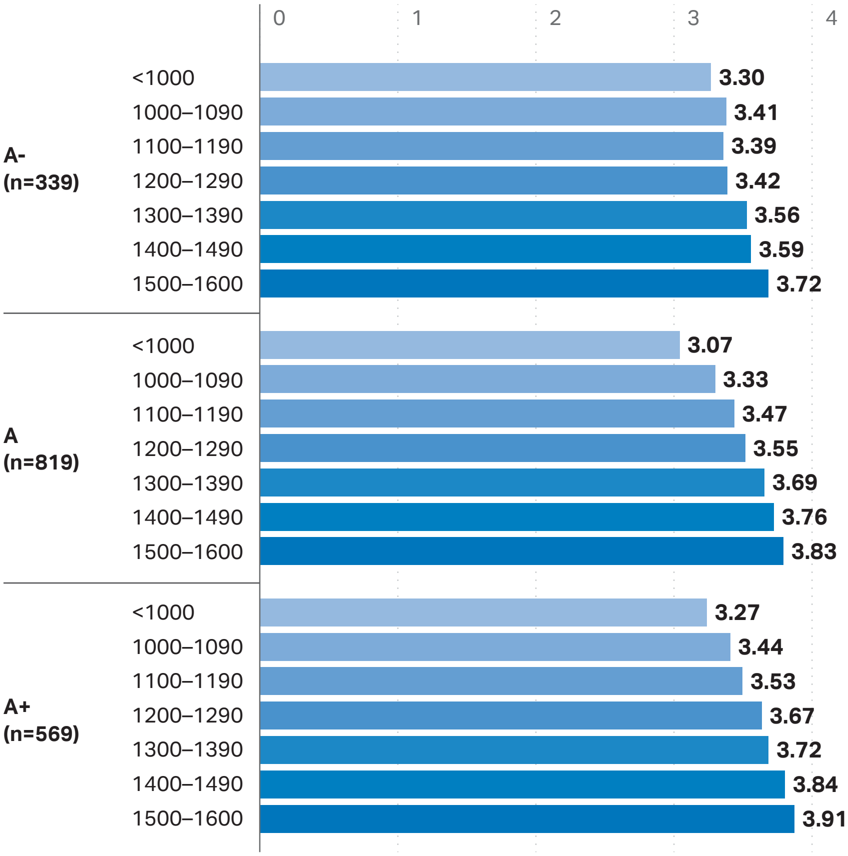
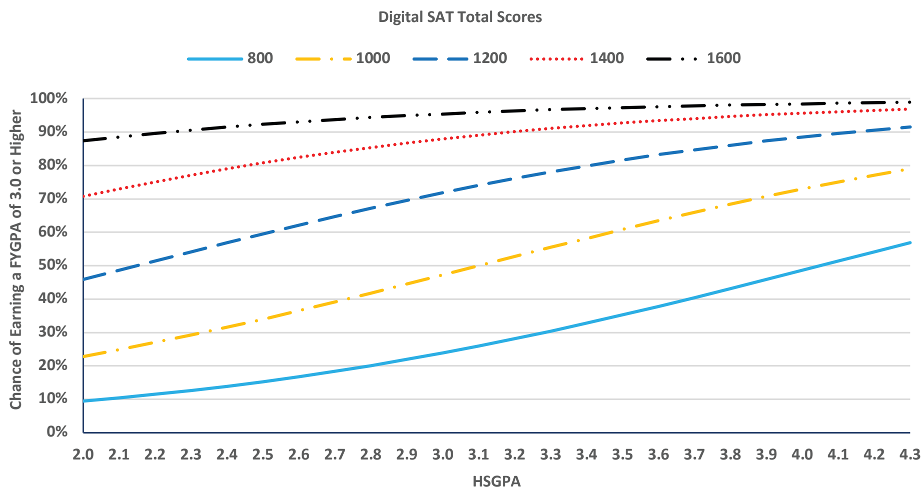



Figure 8.6, addressing the full sample of 1,889 students, shows the estimated probabilities, or chances, of students earning an FYGPA of 3.0 or higher given their digital SAT total score and HSGPA. As with the previously discussed correlational analyses, these logistic regression analyses were conducted at the institution level. Institutional results were then weighted by the number of students, aggregated, and divided by the total number of students across institutions to obtain mean estimates. The figure demonstrates that students with the same HSGPA but different digital SAT total scores have different estimated probabilities of earning an FYGPA of 3.0 or higher. For example, students with an HSGPA of 4.0 and an SAT total score of 1000 have a 73% chance of earning an FYGPA of 3.0 or higher, while students with the same HSGPA but an SAT total score of 1400 have a 96% chance of earning an FYGPA of 3.0 or higher. If SAT total scores did not differentiate college performance among students with the same HSGPA (i.e., add value beyond the information HSGPA alone provides), the colored lines would all be stacked on top of each other; however, this is not the case.

**FIGURE 8.6** PROBABILITY OF EARNING A FIRST-YEAR GPA OF 3.00 OR HIGHER, BY SAT TOTAL SCORE AND HSGPA




The joint use of SAT total scores with HSGPA in a compensatory model such as the one illustrated above helps institutions predict a student's likelihood of succeeding in college even when students perform poorly on either of the two predictors. Using HSGPA alone, conversely, reduces an institution's ability to identify applicants who may perform well academically despite having low high school grades as well as applicants who may face academic difficulties despite having superior high school grades.

## Domain-Specific FYGPAs

### By Digital SAT Section Scores

Figures 8.7 through 8.9 show the positive relationships between digital SAT section scores and domain-specific FYGPAs. Figures 8.7 and 8.8 show the relationships between digital SAT Math section scores and first-year math GPA and first-year STEM GPA, respectively. Across the four Math section score bands—400 to 490, 500 to 590, 600 to 690, and 700 to 800—students' mean math and STEM GPAs increase in stairstep fashion, from 2.92 to 3.60 for first-year math GPA and from 3.01 to 3.68 for first-year STEM GPA. Similarly, as shown in Figure 8.9, first-year all-but-math GPAs rose in tandem with digital SAT Reading and Writing section score bands, from 3.30 to 3.75. Together, these graphs illustrate that digital SAT section scores provide helpful information about likely first-year college performance in the related academic domains and show that digital SAT scores can be useful indicators of readiness for college-level work in particular course areas, thereby productively informing placement decisions.

**FIGURE 8.7** MEAN FIRST-YEAR MATH GPA BY DIGITAL SAT MATH SECTION SCORE BAND

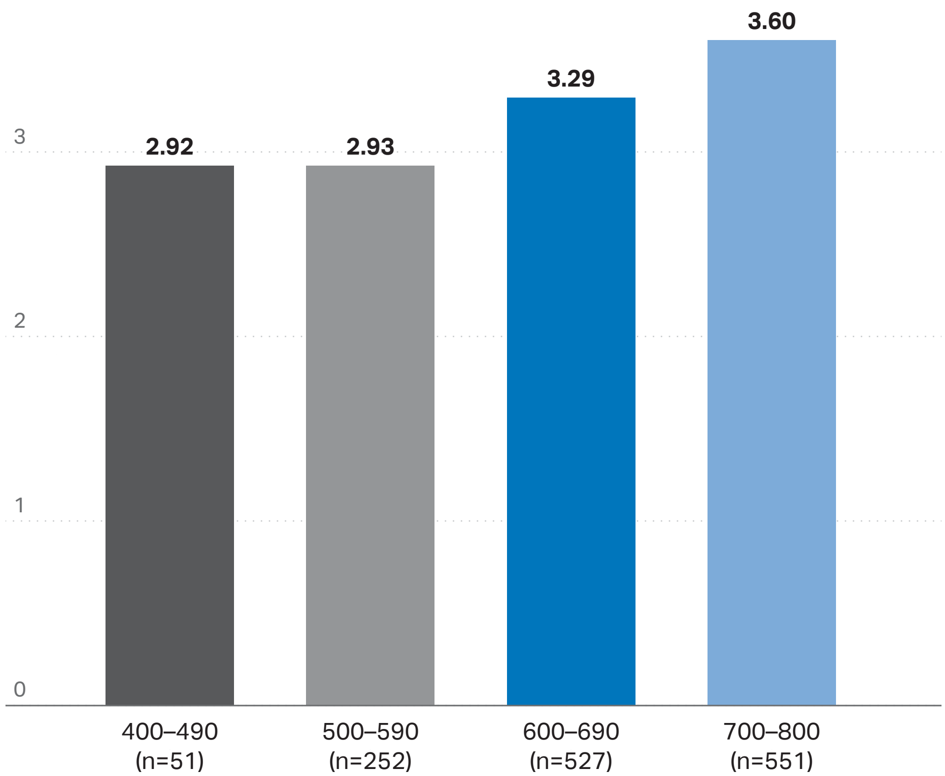



**FIGURE 8.8** MEAN FIRST-YEAR STEM GPA BY DIGITAL SAT MATH SECTION SCORE BAND

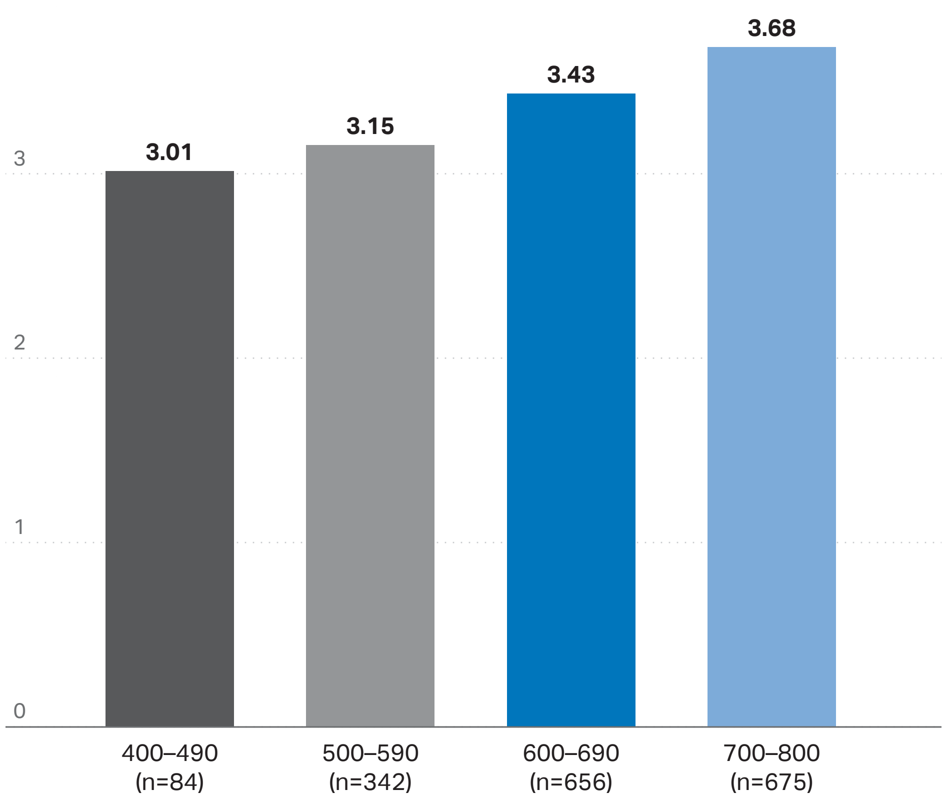
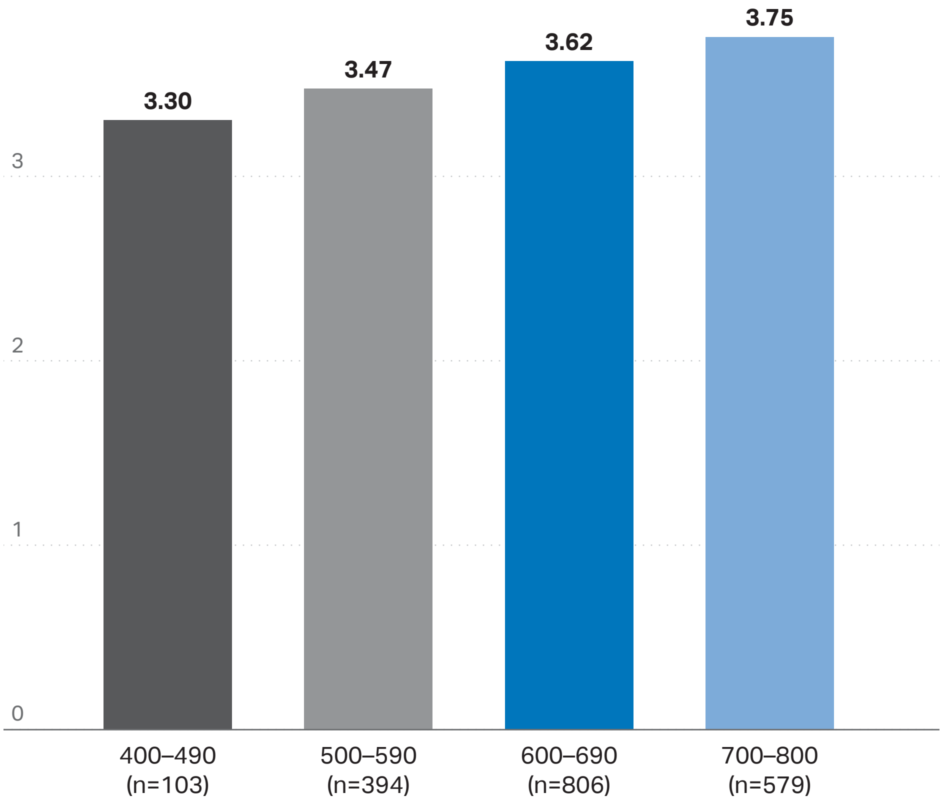


**FIGURE 8.9** MEAN FIRST-YEAR ALL-BUT-MATH GPA BY DIGITAL SAT READING AND WRITING SECTION SCORE BAND




## Controlling for HSGPA

As was done for the earlier FYGPA analyses, College Board researchers conducted logistic regression analyses for the domain-specific GPAs at the institution level. Institutional results were then weighted by the number of students, aggregated, and divided by the total number of students across institutions to obtain mean estimates. Figures 8.10 through 8.12 show students' estimated probabilities, or chances, of earning a domain-specific GPA of 3.0 or higher given a student's corresponding digital SAT section score and HSGPA. At every point along the HSGPA scale for all three domains, students with higher SAT section scores have a higher chance of earning a domain-specific GPA of 3.0 or higher. For example, a student with an SAT Math section score of 700 and a HSGPA of 4.0 has an 87% chance of earning a math FYGPA of 3.0 or higher, while a student with the same HSGPA and an SAT Math section score of 500 has a 49% chance of earning a math FYGPA of 3.0 or higher. These figures clearly indicate the informational value added by SAT section scores, over and above the use of HSGPA alone, in predicting how students will perform in particular academic domains in college. Institutions may choose to run similar logistic regression analyses at their institutions in order to use SAT section scores (and HSGPA) to inform course placement decisions.

**FIGURE 8.10** PROBABILITY OF EARNING A MATH FYGPA OF 3.00 OR HIGHER, BY DIGITAL SAT MATH SECTION SCORE AND HSGPA

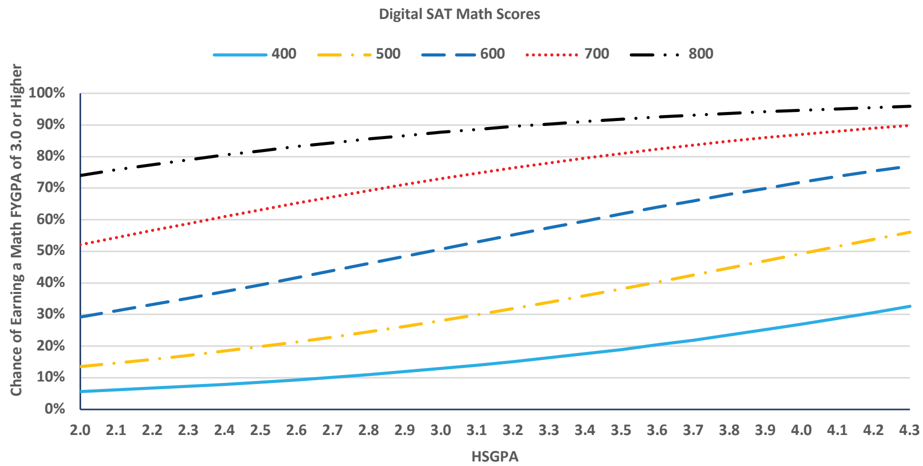



**FIGURE 8.11** PROBABILITY OF EARNING A STEM FYGPA OF 3.00 OR HIGHER, BY DIGITAL SAT MATH SECTION SCORE AND HSGPA

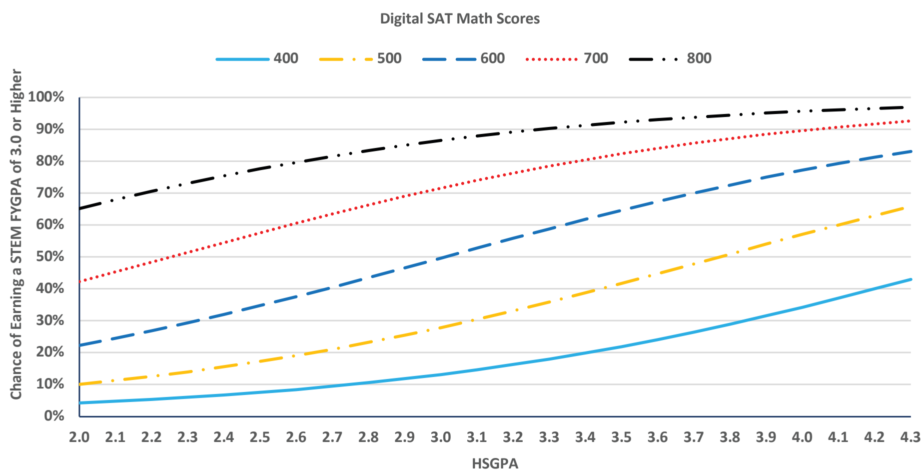
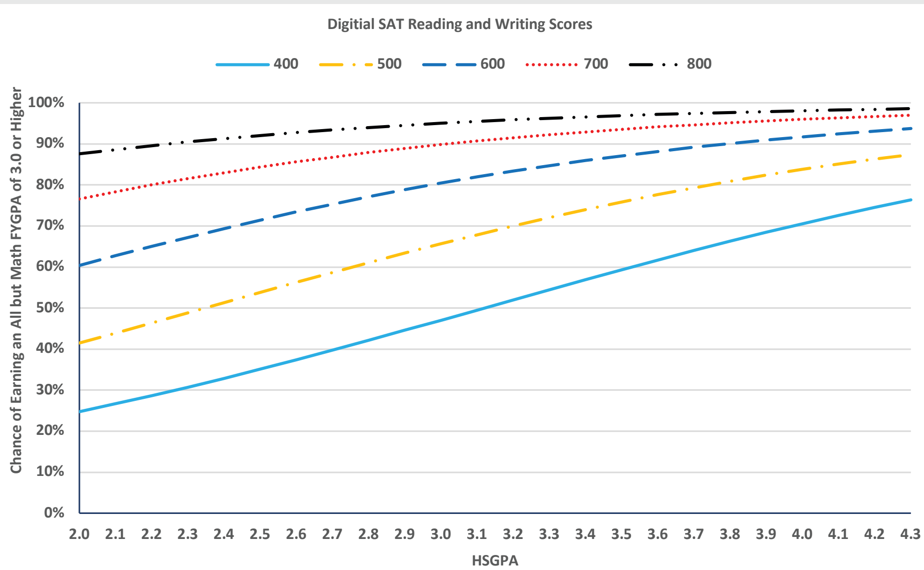


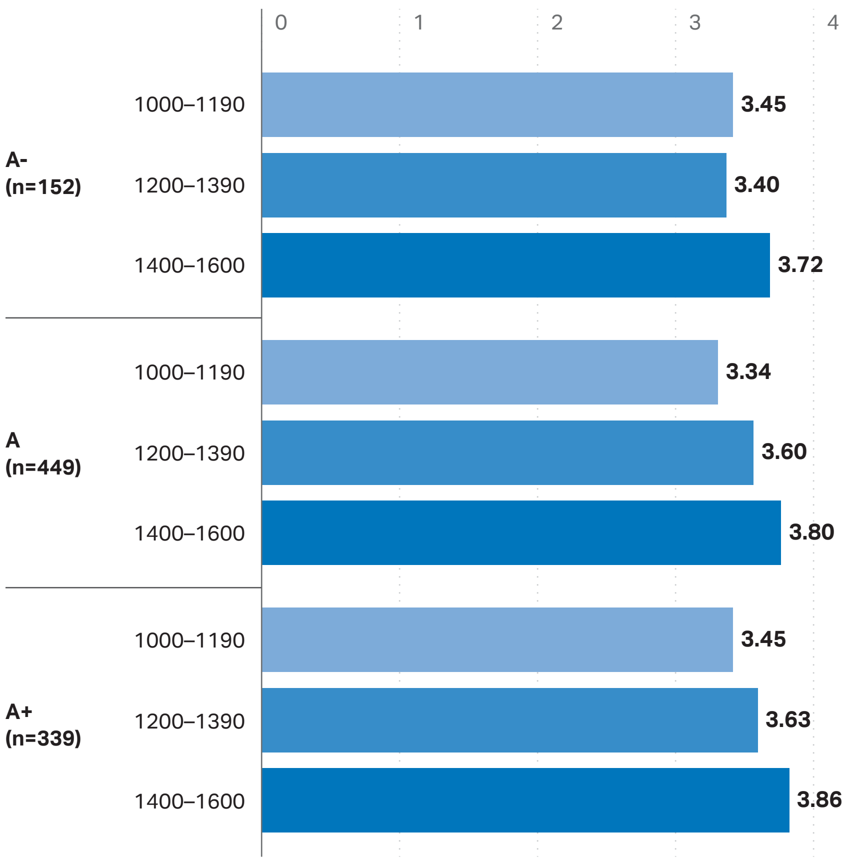

**FIGURE 8.12** PROBABILITY OF EARNING AN ALL-BUT-MATH FYGPA OF 3.00 OR HIGHER, BY DIGITAL SAT READING AND WRITING SECTION SCORE AND HSGPA



## STEM Majors

In addition to examining digital SAT score relationships with domain-specific GPAs, College Board researchers analyzed digital SAT score relationships with FYGPA for students majoring in STEM fields in their first year of college. The researchers examined the added value of using digital SAT total scores in addition to HSGPA to understand students' overall first-year academic performance. This can be helpful in considering performance for future students. Figure 8.13 illustrates not only the strong, positive relationship between digital SAT total scores and FYGPA for STEM majors but also the incremental utility gained by using digital SAT total scores alongside HSGPA in predictions of FYGPA. Among STEM majors, correlations with FYGPA were found to be .64 for the digital SAT total score, .52 for HSGPA, and .72 for the digital SAT total score used in conjunction with HSGPA, the last being an improvement of 38% over the use of HSGPA alone.[19] This demonstrates the large contribution that digital SAT score information can make to understanding how STEM majors will likely perform in the first year of college.

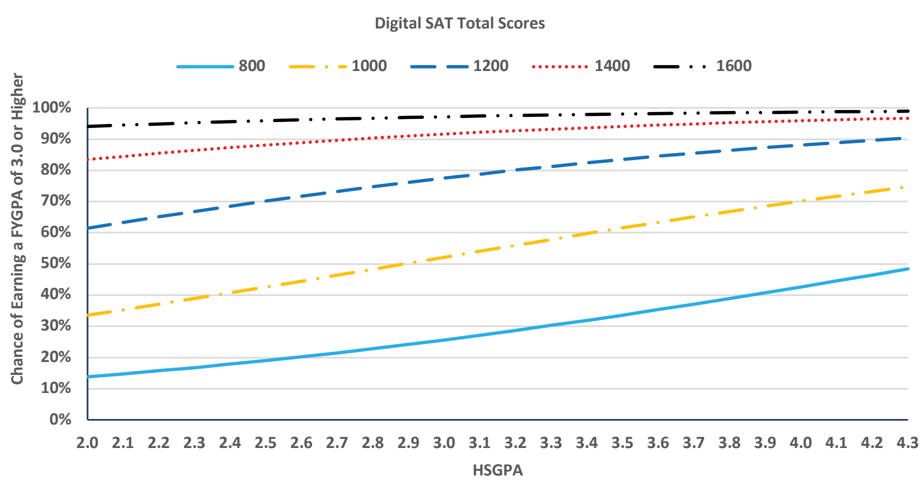
**FIGURE 8.13** MEAN FIRST-YEAR GPA FOR STEM MAJORS, BY DIGITAL SAT TOTAL SCORE BAND WITHIN HSGPA




19 This value was calculated by subtracting the HSGPA-FYGPA correlation (.52) from the multiple correlation of HSGPA and SAT with FYGPA (.72) to arrive at the SAT incremental validity coefficient (.20). This coefficient is then divided by the HSGPA-FYGPA correlation (.52) and multiplied by 100 to arrive at the incremental predictive utility value of 38%.

Figure 8.14 shows STEM majors' estimated probabilities, or chances, of earning an FYGPA of 3.0 or higher given their digital SAT total score and HSGPA. Once again, students with the same HSGPA but different SAT total scores have different estimated probabilities of earning an FYGPA of 3.0 or higher. For example, STEM majors with an HSGPA of 4.0 and an SAT total score of 1000 have a 70% chance of earning an FYGPA of 3.0 or higher, while students with the same HSGPA and an SAT total score of 1400 have a 99% chance of earning an FYGPA of 3.0 or higher. This figure demonstrates how SAT total scores, used in conjunction with HSGPA, can quickly and productively inform decisions about which students may be successful in competitive academic majors and which students may need additional academic support to be successful in those majors.

**FIGURE 8.14** PROBABILITY OF STEM MAJORS EARNING A STEM FYGPA OF 3.00 OR HIGHER, BY DIGITAL SAT TOTAL SCORE AND HSGPA



## First-Year Credits Earned

In addition to FYGPA, College Board researchers examined relationships between digital SAT scores and credits earned in the first year of college. Given that most bachelor's degree programs require that students earn 120 credits, students completing 30 credits in their first year of college are on track to graduate within four years. As shown in Table 8.13, digital SAT scores and HSGPA individually have moderately strong, positive relationships with first-year credits earned, with correlations of .45 and .39, respectively; when used jointly, the relationship is stronger, with a multiple correlation of .50. This represents a 28% increase in predictive utility over using HSGPA alone.

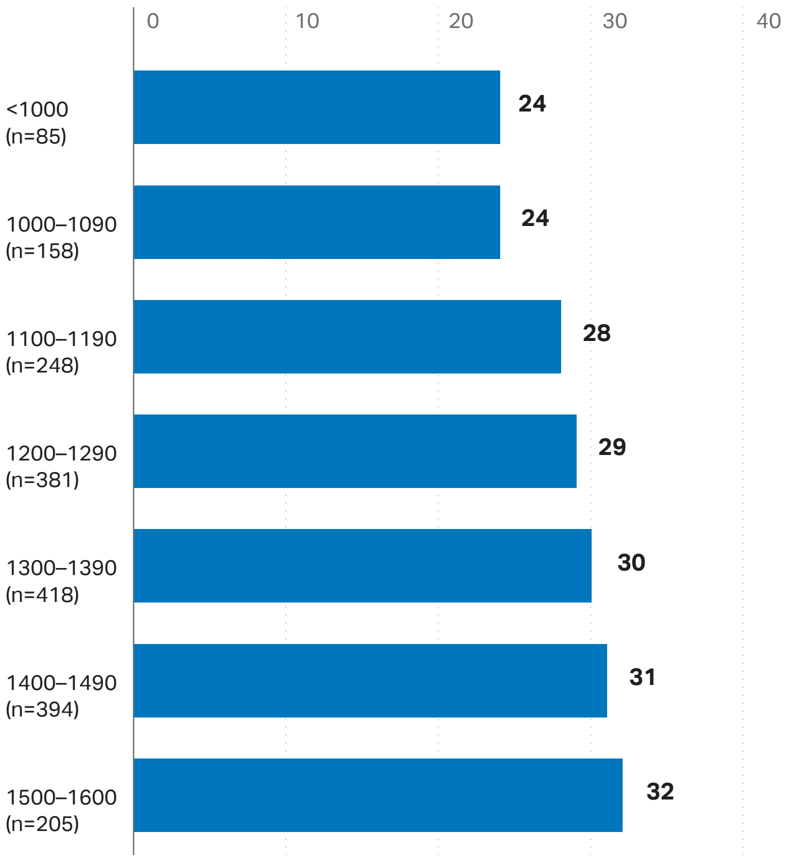
**TABLE 8.13 CORRECTED (RAW) CORRELATION OF PREDICTORS WITH FIRST-YEAR CREDITS EARNED (K=11, N=1,889)**

| Predictor(s) | Correlation | 95% Confidence Interval |
|---|---|---|
| Digital SAT Reading and Writing Section Score | .39 (.21) | .35–.43 |
| Digital SAT Math Section Score | .43 (.28) | .39–.47 |
| Digital SAT Total Score | .45 (.30) | .41–.49 |
| HSGPA | .39 (.17) | .35–.43 |
| Digital SAT+HSGPA | .50 (.33) | .47–.53 |
| Digital SAT incremental validity beyond the use of HSGPA alone | .11 (.16) | |

*Note.* Confidence intervals were calculated using the adjusted correlations after rounding.

Figure 8.15 illustrates the positive relationship between digital SAT total scores and first-year credits earned. As digital SAT total scores increase, the mean number of credits earned[20] also increases in a stairstep manner.

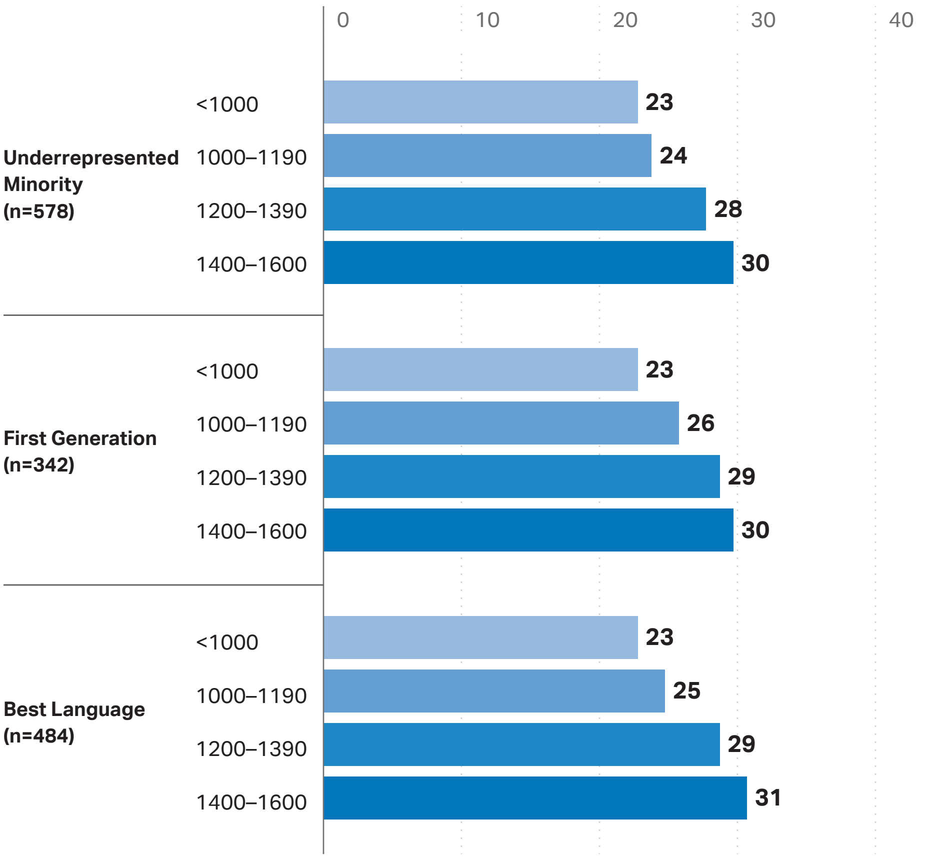
**FIGURE 8.15** MEAN FIRST-YEAR CREDITS BY DIGITAL SAT TOTAL SCORE BAND




20 These are credits earned on campus. AP credits are not included.

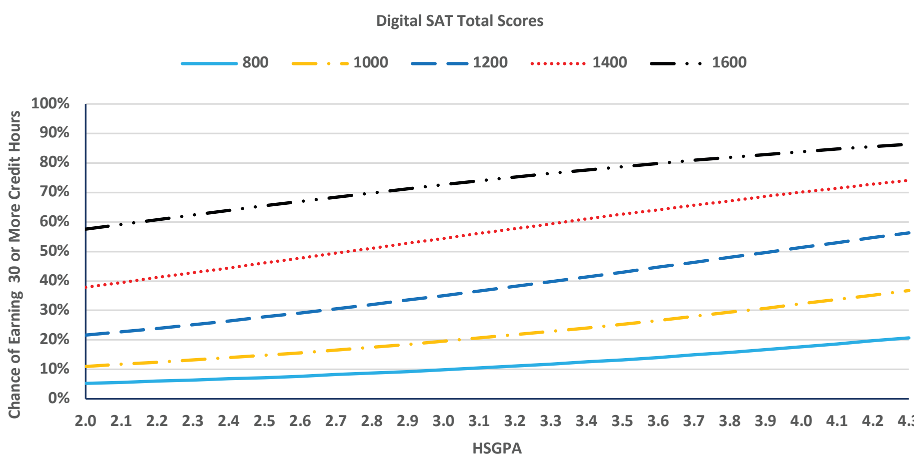
As with the FYGPA analyses, College Board researchers examined the relationship between digital SAT total scores and credits earned in the first year of college by underrepresented minority students, first-generation students, and students whose self-identified best language is English and another language or a language other than English. The results depicted in Figure 8.16 indicate a clear, positive relationship between digital SAT total scores and credits earned by all three subgroups, with credits earned rising in tandem with digital SAT total scores.

**FIGURE 8.16** **MEAN FIRST-YEAR CREDITS EARNED BY DIGITAL SAT TOTAL SCORE BAND: SUBGROUP ANALYSES**



Lastly, Figure 8.17 presents the results of logistic regression analyses in which College Board researchers used digital SAT total scores in conjunction with HSGPA to determine students' probabilities of earning 30 or more credits in their first year of college. For students with the same HSGPA, their chances of earning 30 or more credits in their first year vary according to their digital SAT total scores. For example, for students in this study with a HSGPA of 4.0 and a digital SAT total score of 1600, 1400, 1200, 1000, or 800, their estimated chances of earning 30 or more credits were 84%, 70%, 51%, 32%, and 18%, respectively. Without using digital SAT scores as part of the calculation, the estimated chances for these students would be identical. These findings demonstrate the value of using digital SAT total scores in conjunction with HSGPA rather than using HSGPA alone to identify students who may struggle to complete a bachelor's degree within four years.

**FIGURE 8.17** PROBABILITY OF EARNING 30 OR MORE CREDITS IN THE FIRST YEAR OF COLLEGE, BY DIGITAL SAT TOTAL SCORE AND HSGPA




## 8.5.5 Summary

Collectively, the study results presented in this section show that digital SAT scores are as predictive of college performance as are paper and pencil SAT scores and that their use meaningfully improves the ability to predict college performance above the use of HSGPA alone. Strong positive relationships between digital SAT scores and FYGPA were also documented for select population subgroups, such as underrepresented minority students, first-generation college students, and students whose self-identified best language(s) were English and another language or a language other than English.

When performance in specific first-year coursework domains was examined, strong relationships were observed between digital SAT Math section scores and both math and STEM course grades as well as between digital SAT Reading and Writing section scores and students' performance in courses other than math. When digital SAT scores were examined for students majoring in STEM fields, even stronger relationships were observed than for the overall sample, with a 38% improvement in the prediction of college performance from the use of digital SAT scores over and above the use of HSGPA alone.

The current study also found a positive relationship between digital SAT scores and the number of first-year college credits earned, a proxy for understanding progress toward degree completion. These findings suggest that students with higher digital SAT scores tend not only to earn higher grades but also to make quicker progress toward completing a bachelor's degree.

In sum, these findings should give institutions confidence that digital SAT scores provide valuable insights about students' readiness for college; can productively inform course placement, academic major, and scholarship and honors program decisions; and can help identify students who may need academic support. In the years following the completion of the transition to digital testing in 2024, College Board will study the first entering college cohort with digital SAT scores to longitudinally examine digital SAT score relationships with college outcomes across a large, representative national sample of students and institutions.

## 8.6 Measuring and Monitoring College and Career Readiness with the Digital SAT Suite

The digital SAT Suite College and Career Readiness Benchmarks and grade-level benchmarks help students, families, and educators assess student progress toward readiness for college and workforce training from year to year. These benchmarks help:

- Identify students who are thriving and require greater challenges
- Identify students who require additional academic support
- Inform instructional and curricular enhancements throughout an institution

### College and Career Readiness Benchmarks

Students are considered college and career ready when both their digital SAT Reading and Writing and Math section scores meet established benchmarks. It is important to note that college and career readiness operates on a continuum and is not a dichotomous, either/or condition: students scoring below the digital SAT College and Career Readiness Benchmarks can still be successful in college, especially with additional preparation and perseverance.

The digital SAT College and Career Readiness benchmarks were empirically derived from analysis of large, representative samples of student performance over many years. They represent the minimum scores on the Reading and Writing and Math sections that are associated with having a high likelihood—specifically, a 75% probability—of earning at least a C in relevant credit-bearing, introductory college-level courses. The Math College and Career Readiness Benchmark score of 530, for example, is the minimum section score needed for a student to have a 75% chance of obtaining at least a C in first-semester, credit-bearing college-level courses in college algebra, statistics, precalculus, and/or calculus. If a student completed more than one relevant course in the first semester, the course with the lowest grade is used. Similarly, the Reading and Writing College and Career Readiness Benchmark score of 480 is the minimum section score a student can earn to have a 75% chance of obtaining at least a C in first-semester, credit-bearing college-level courses in history, literature, social science, and/or writing. If a student completed more than one relevant course in the first semester, the course with the lowest grade is used. Using the lowest course grade for both Reading and Writing and Math ensures that students who meet or exceed the benchmark scores are prepared to succeed in all their first-semester, introductory credit-bearing college courses. Validation analyses have indicated that students at two- and four-year postsecondary institutions meeting both the Reading and Writing and Math College and Career Readiness benchmarks had higher retention rates, higher grades in college, and graduated at higher rates than those who did not meet the benchmarks (College Board, 2017).

## Grade-Level Benchmarks

Students who take the PSAT/NMSQT, PSAT 10, and PSAT 8/9 receive grade-level benchmark information on their score reports. Grade-level benchmarks indicate whether students are on track for college and career readiness by the end of high school and are based on expected student growth at each grade toward the College and Career Readiness Benchmarks. Average performance on the Reading and Writing and Math sections was computed for each grade year in the benchmark analysis datasets, and then estimated average yearly growth estimates were calculated by subtracting the average performance in each year from the average performance in the subsequent year. Grade-level benchmarks were then computed by subtracting the average growth from the subsequent year's benchmark. For detailed information on how the grade-level benchmarks were calculated, see the 2017 *SAT Suite of Assessments Technical Manual* (College Board, 2017).

## Benchmark Scores

Table 8.14 displays the digital SAT Suite benchmark scores by test section.

**TABLE 8.14 DIGITAL SAT SUITE COLLEGE AND CAREER READINESS AND GRADE-LEVEL BENCHMARK SCORES**

| Benchmark | Reading and Writing Section Benchmark Score | Math Section Benchmark Score |
|---|---|---|
| College and Career Readiness | 480 | 530 |
| Grade 11 | 460 | 510 |
| Grade 10 | 430 | 480 |
| Grade 9 | 410 | 450 |
| Grade 8 | 390 | 430 |

## Future Analyses

At the time of this writing, College Board expects the digital-suite College and Career Readiness benchmarks to remain the same as those for the paper and pencil suite, and the organization will conduct additional research to determine whether any small changes should be made to the grade-level benchmarks. The benchmarks are generally expected to remain the same because of the straight-line concordance produced for all test scores on the SAT, as well as the fact that College Board was able to maintain the SAT Suite's vertical scale in the transition to the digital exams. As the organization gathers actual college performance data for students who have taken the digital SAT Suite tests, all benchmark scores will be reexamined and either validated or updated as needed. This work is expected to occur in 2027. That time frame is also far enough removed to allow College Board to better assess whether now-emerging modifications to college admission policies, shifts in college grading practices and standards, the effects of pandemic learning loss, and/or changes to the digital SAT Suite test-taking population have impacted the placement of digital SAT Suite benchmarks.

Chapter 9

# Conclusion

This manual has made an effort to provide all interested stakeholders of the digital SAT Suite of Assessments with information about the technical qualities of the suite, including the content of the assessments; the procedures and processes that are undertaken in the creation, administration, and scoring of the assessments; how to properly interpret digital SAT Suite scores; the accuracy of digital-suite scores from a measurement perspective; and evidence with bearing on the validity of interpretations made on the basis of the scores. In addition to serving as a resource for stakeholders of the tests, it is also an important component in codifying and maintaining best assessment practices. It is with this in mind that, in addition to the expected technical information, this manual offers insights into some of the changes brought about by the transition to digital testing, including the rationale behind those changes and the benefits they bring to test takers.

Because test taking is a continuously evolving and iterative process, this manual will be updated periodically in order to provide stakeholders with the fullest and most accurate information available.

# 9.0 References


Abedi, J., & Sato, E. (2008). *Linguistic modification*. U.S. Department of Education. **https://ncela.ed.gov/files/uploads/11/abedi_sato.pdf**

ACT. (2006). *Reading between the lines: What the ACT reveals about college readiness in reading*. ACT. **https://files.eric.ed.gov/fulltext/ED490828.pdf**

ACT. (2016). *National curriculum survey 2016*. ACT. **https://www.act.org/content/dam/act/unsecured/documents/NCS_Report_Web_2016.pdf**

ACT. (2018). *The condition of college and career readiness: National 2018*. ACT. **https://www.act.org/content/dam/act/unsecured/documents/cccr2018/National-CCCR-2018.pdf**

ACT. (2020). *ACT national curriculum survey 2020*. ACT. **https://www.act.org/content/dam/act/unsecured/documents/NCS_Report_Web_2020.pdf**

ACT. (2023). *The ACT profile report—national: Graduating class of 2023*. ACT. **https://www.act.org/content/dam/act/unsecured/documents/2023-National-ACT-Profile-Report.pdf**

Adams, M. J. (2009). The challenge of advanced texts. In E. H. Hiebert (Ed.), *Reading more, reading better* (pp. 163–89). Guilford. **https://www.childrenofthecode.org/library/MJA-ChallengeofAdvancedTexts.pdf**

Adelman, C. (2006). *The toolbox revisited: Paths to degree completion from high school through college*. U.S. Department of Education. **https://www2.ed.gov/rschstat/research/pubs/toolboxrevisit/toolbox.pdf**

Ahmed, Y., Francis, D. J., York, M., Fletcher, J. M., Barnes, M., & Kulesz, P. A. (2016). Validation of the Direct and Inferential Mediation (DIME) model of reading comprehension in grades 7 through 12. *Contemporary Educational Psychology*, *44–45*, 68–82. **https://doi.org/10.1016/j.cedpsych.2016.02.002**

American Diploma Project. (2004). *Ready or not: Creating a high school diploma that counts*. Achieve. **https://eric.ed.gov/?id=ED494733**

American Educational Research Association (AERA), American Psychological Association (APA), & National Council on Measurement in Education (NCME). (2014). *Standards for educational and psychological testing*. AERA. **https://www.testingstandards.net/uploads/7/6/6/4/76643089/standards_2014edition.pdf**

American Statistical Association. (n.d.). *College majors requiring statistics*. American Statistical Association. **https://www.amstat.org/asa/files/pdfs/EDU-CollegeMajorsFlyer.pdf**

Andrich, D. (1988). *Rasch models for measurement*. SAGE.

Asim, M., Kurlaender, M., & Reed, S. (2019). *12th grade course-taking and the distribution of opportunity for college readiness in mathematics*. Policy Analysis for California Education. **https://files.eric.ed.gov/fulltext/ED600439.pdf**

Atuahene, F., & Russell, T. A. (2016). Mathematics readiness of first-year university students. *Journal of Developmental Education*, *39*(3), 12–32. **https://files.eric.ed.gov/fulltext/EJ1130188.pdf**

Bautsch, B. (2013). *Reforming remedial education*. Hot Topics in Higher Education, National Conference of State Legislatures. **https://dcmathpathways.org/sites/default/files/resources/2017-03/State%20Policies%20to%20Improve%20Student%20Transfer.pdf**

Beason, L. (2001). Ethos and error: How business people react to errors. *College Composition and Communication*, *53*(1), 33–64. **https://doi.org/10.2307/359061**

Beck, I. L., McKeown, M. G., & Kucan, L. (2013). *Bringing words to life: Robust vocabulary instruction* (2nd ed.). Guilford.

Beers, G. K., & Probst, R. E. (2012). *Notice and note: Strategies for close reading.* Heinemann.

Biag, M., & Williams, I. (2014). *Examining students' math course-taking and educational outcomes in the Garden Grove Unified School District (year 1).* John W. Gardner Center for Youth and Their Communities. **https://files.eric.ed.gov/fulltext/ED573648.pdf**

Biemiller, A. (2005). Size and sequence in vocabulary development: Implications for choosing words for primary grade vocabulary instruction. In E. H. Hiebert & M. L. Kamil (Eds.), *Teaching and learning vocabulary: Bringing research to practice* (pp. 223–42). Routledge.

Blair, R., Kirkman, E. E., & Maxwell, J. W. (2018). *Statistical abstract of undergraduate programs in the mathematical sciences in the United States: Fall 2015 CBMS survey.* American Mathematical Society. **http://www.ams.org/profession/data/cbms-survey/cbms2015-Report.pdf**

Britton, B. K., & Graesser, A. C. (Eds.). (2014). *Models of understanding text.* Psychology Press. **https://doi.org/10.4324/9781315806143**

Brumberger, E. (2011). Visual literacy and the digital native: An examination of the millennial learner. *Journal of Visual Literacy*, *30*(1), 19–47. **https://doi.org/10.1080/23796529.2011.11674683**

Bruno, P., & Goldhaber, D. (2021). What pandemic-related test waiver requests suggest about states' testing priorities. *Phi Delta Kappan*, *103*(3), 48–53. **https://doi.org/10.1177/00317217211058525**

Brysbaert, M., & Biemiller, A. (2017). Text-based age-of-acquisition norms for 44 thousand English word meanings. *Behavior Research Methods*, *49*(4), 1520–23. **https://doi.org/10.3758/s13428-016-0811-4**

Cai, L. (2017). flexMIRT version 3.51: Flexible multilevel multidimensional item analysis and test scoring [Computer software]. Vector Psychometric Group.

Cain, K., & Oakhill, J. V. (1999). Inference making ability and its relation to comprehension failure in young children. *Reading and Writing*, *11*(5–6), 489–503. **https://doi.org/10.1023/A**:1008084120205

Carnevale, A. P., & Fasules, M. L. (2021, March 9). *How useful is high school math on the job?* Medium. **https://medium.com/georgetown-cew/how-useful-is-high-school-math-on-the-job-f5ae9002df12**

Carpenter, T. P., Corbitt, M. K., Kepner Jr., H. S., Lindquist, M. M., & Reys, R. (1980). Results of the second NAEP mathematics assessment: Secondary school. *The Mathematics Teacher*, *73*(5), 329–38. **https://www.jstor.org/stable/27962023**

Casteel, M. A., & Simpson, G. B. (1991). Textual coherence and the development of inferential generation skills. *Journal of Research in Reading*, *14*(2), 116–29. **https://doi.org/10.1111/j.1467-9817.1991.tb00013.x**

Center for Universal Design. (1997, April 1). *The principles of universal design* (version 2.0). North Carolina State University, Center for Universal Design. **https://design.ncsu.edu/research/center-for-universal-design**

Cervetti, G. N., Wright, T. S., & Hwang, H. (2016). Conceptual coherence, comprehension, and vocabulary acquisition: A knowledge effect? *Reading and Writing*, *29*(4), 761–79. **https://doi.org/10.1007/s11145-016-9628-x**

Chall, J. S., & Jacobs, V. A. (2003). The classic study on poor children's fourth-grade slump. *American Educator*, *27*(1), 14–15. **https://www.aft.org/periodical/american-educator/spring-2003/classic-study-poor-childrens-fourth-grade-slump**

Citizens Financial Group. (2021, August 4). *Citizens survey shows that affordability concerns remain top of mind for students and families*. Business Wire. **https://www.businesswire.com/news/home/20210804005179/en**

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Education and Psychological Measurement*, *20*(1), 37–46. **https://doi.org/10.1177/001316446002000104**

Cohen, J. (1968). Weighted kappa: Nominal scales agreement provision for scaled disagreement or partial credit. *Psychological Bulletin*, *70*, 213–20. **https://doi.org/10.1037/h0026256**

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

College Board. (2017). *SAT Suite of Assessments technical manual: Characteristics of the SAT*. College Board. **https://satsuite.collegeboard.org/media/pdf/sat-suite-assessments-technical-manual.pdf**

College Board. (2018). *Guidelines on the uses of College Board test scores and related data*. College Board. **https://research.collegeboard.org/pdf/guidelines-uses-college-board-test-scores-and-data.pdf**

College Board. (2019). *College Board national curriculum survey report 2019*. College Board. **https://satsuite.collegeboard.org/media/pdf/national-curriculum-survey-report.pdf**

College Board. (2022). *The digital SAT Suite of Assessments specifications overview*. College Board. **https://satsuite.collegeboard.org/media/pdf/digital-sat-test-spec-overview.pdf**

College Board. (2023a). *2023 SAT Suite of Assessments annual report: Total group*. College Board. **https://reports.collegeboard.org/media/pdf/2023-total-group-sat-suite-of-assessments-annual-report%20ADA.pdf**

College Board. (2023b). *Trends in college pricing and student aid 2023*. College Board. **https://research.collegeboard.org/media/pdf/Trends%20Report%202023%20Updated.pdf**

College Board. (2024). *The cognitively complex thinking required by select digital SAT Suite questions*. College Board. **https://satsuite.collegeboard.org/media/pdf/digital-sat-cognitive-lab-report.pdf**

College Board & HumRRO. (2020). *The complex thinking required by select SAT items: Evidence from student cognitive interviews*. College Board. **https://satsuite.collegeboard.org/media/pdf/sat-cognitive-lab-report.pdf**

Cunningham, A. E., & Stanovich, K. E. (1998). What reading does for the mind. *American Educator*, *22*(1–2), 8–15. **https://eric.ed.gov/?id=EJ571299**

Dale, E., & O'Rourke, J. (1981). *The living word vocabulary: The words we know, a national vocabulary inventory*. World Book–Childcraft International.

Dong, Y., Peng, S., Sun, Y., Wu, S. X., & Wang, W. (2020). Reading comprehension and metalinguistic knowledge in Chinese readers: A meta-analysis. *Frontiers in Psychology*, *10*, Article 3037, 1–15. **https://doi.org/10.3389/fpsyg.2019.03037**

Dorans, N. J., & Feigenbaum M. D. (1994). Equating issues engendered by changes to the SAT and PSAT/NMSQT. In I. M. Lawrence, N. J. Dorans, M. D. Feigenbaum, N. J. Feryok, A. P. Schmitt, & N. K. Wright (Eds.), *Technical issues related to the introduction of the new SAT and PSAT/NMSQT*. Research memorandum no. RM-94-10, Educational Testing Service (ETS). **https://www.ets.org/research/policy_research_reports/publications/report/1994/ichq.html**

Drasgow, F., Luecht, R. M., & Bennett, R. E. (2006). Technology and Testing. In R. L. Brennan (Ed.), *Educational measurement*, 4th ed. (pp. 471–515). American Council on Education/Praeger.

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest*, *14*(1), 4–58. **https://doi.org/10.1177/1529100612453266**

Engel, J. (2017). Statistical literacy for active citizenship: A call for data science education. *Statistics Education Research Journal*, *16*(1), 44–49. **http://iase-web.org/documents/SERJ/SERJ16(1)_Engel.pdf**

Er, S. N. (2018). Mathematics readiness of first-year college students and missing necessary skills: Perspectives of mathematics faculty. *Journal of Further and Higher Education*, *42*(7), 937–952. **https://doi.org/10.1080/0309877X.2017.1332354**

Fey, J. T. (Ed.). (1984). *Computing and mathematics: The impact on secondary school curricula*. National Council of Teachers of Mathematics.

Fisher, D., Frey, N., Anderson, H., & Thayre, M. (2014). *Text-dependent questions, grades 6–12: Pathways to close and critical reading*. Corwin.

Fishman, R., Johnson, E., Nguyen, S., & Romo-González, L. (2021). *Varying Degrees 2021: New America's fifth annual survey on higher education*. New America. **https://www.newamerica.org/education-policy/reports/varying-degrees-2021**

Freudenthal, H. (1971). Geometry between the devil and the deep sea. *Educational Studies in Mathematics*, *3*(3/4), 413–35. **https://www.jstor.org/stable/3482035**

Fry, E. (1981). Graphical literacy. *Journal of Reading 24*(5), 383–89. **https://www.jstor.org/stable/40032373**

Fu, Y., Choe, E. M., Lim, H., & Choi, J. (2022). An evaluation of automatic item generation: A case study of weak theory approach. *Educational Measurement: Issues and Practice*, *41*(4), 1–13. **https://doi.org/10.1111/emip.12529**

GAISE College Report ASA Revision Committee. (2016). *Guidelines for Assessment and Instruction in Statistics Education (GAISE) college report 2016*. American Statistical Association. **https://www.amstat.org/asa/files/pdfs/GAISE/GaiseCollege_Full.pdf**

Galesic, M., & Garcia-Retamero, R. (2011). Graph literacy: A cross-cultural comparison. *Medical Decision Making*, *31*(3), 444–57. **https://journals.sagepub.com/doi/10.1177/0272989X10373805**

Gallup & Lumina Foundation. (2015). *Postsecondary education aspirations and barriers: The 2014 Gallup–Lumina Foundation study of the American public's opinion on higher education*. Gallup. **https://www.luminafoundation.org/resource/postsecondary-education-aspirations-and-barriers**

Gamoran, A., & Hannigan, E. C. (2000). Algebra for everyone? Benefits of college-preparatory mathematics for students with diverse abilities in early secondary school. *Educational Evaluation and Policy Analysis*, *22*(3), 241–54. **https://doi.org/10.3102/01623737022003241**

Gierl, M. J., Zhou, J., & Alves, C. (2008). Developing a taxonomy of item model types to promote assessment engineering. *Journal of Technology, Learning, and Assessment*, *7*(2), 4–50. **https://files.eric.ed.gov/fulltext/EJ838625.pdf**

Ginsburg, A., Cooke, G., Leinwand, S., Noell, J., & Pollock, E. (2005). *Reassessing U.S. international mathematics performance: New findings from the 2003 TIMSS and PISA*. American Institutes for Research. **https://files.eric.ed.gov/fulltext/ED491624.pdf**

Gottfried, M. A., Bozick, R., & Srinivasan, S. V. (2014). Beyond academic math: The role of applied STEM course taking in high school. *Teachers College Record*, *116*(7), 1–35. **http://dx.doi.org/10.1177/016146811411600703**

Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review*, *101*(3), 371–95. **https://doi.org/10.1037/0033-295X.101.3.371**

Hallgren, K. A. (2012). Computing interrater reliability for observational data: An overview and tutorial. *Tutorials in Quantitative Methods for Psychology*, *8*(1), 23. **https://www.tqmp.org/RegularArticles/vol08-1/p023**

Han, K. T. (2007). WinGen version 3.1: Windows software that generates IRT parameters and item responses [Computer software]. University of Massachusetts

Hanson, M. (2023a, July 16). *Student loan debt by gender*. Education Data Initiative. **https://educationdata.org/student-loan-debt-by-gender**

Hanson, M. (2023b, December 8). *Student loan debt by race*. Education Data Initiative. **https://educationdata.org/student-loan-debt-by-race**

Hanson, M. (2024, March 3). *Student loan debt statistics*. Education Data Initiative. **https://educationdata.org/student-loan-debt-statistics**

Harsh, J. A., & Schmitt-Harsh, M. (2016). Instructional strategies to develop graphing skills in the college science classroom. *The American Biology Teacher*, *78*(1), 49–56. **https://doi.org/10.1525/abt.2016.78.1.49**

Hart Research Associates. (2018). *Fulfilling the American dream: Liberal education and the future of work—selected findings from online surveys of business executives and hiring managers*. Association of American Colleges and Universities. **https://www.aacu.org/research/fulfilling-the-american-dream-liberal-education-and-the-future-of-work**

Heath, S. B. (1983). *Ways with words: Language, life, and work in communities and classrooms*. Cambridge University Press.

Helms, J. (2015, February 25). *Five examples of how federal agencies use big data*. IBM Center for the Business of Government [Blog]. **https://www.businessofgovernment.org/blog/five-examples-how-federal-agencies-use-big-data**

Hendrickson, A. (2007). An NCME instructional module on multistage testing. *Educational Measurement: Issues and Practice*, *26*(2), 44–52. **https://doi.org/10.1111/j.1745-3992.2007.00093.x**

Houts, C. R., & Cai, L. (2016). flexMIRT user's manual version 3.5: Flexible multilevel multidimensional item analysis and test scoring. Vector Psychometric Group.

Intersegmental Committee of the Academic Senates (ICAS) of the California Community Colleges, the California State University, and the University of California. (2002). *Academic literacy: A statement of competencies expected of students entering California's public colleges and universities*. ICAS. **https://senate.universityofcalifornia.edu/_files/reports/acadlit.pdf**

Kane, M. T. (1992). An argument-based approach to validity. *Psychological Bulletin*, *112*(3), 527–35. **https://doi.org/10.1037/0033-2909.112.3.527**

Kane, M. T. (2006). Validation. In R. Brennan (Ed.), *Educational measurement*, 4th ed. (pp. 17–64). American Council on Education/Praeger.

Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement*, *50*, 1–73. **https://doi.org/10.1111/jedm.12000**

Kaplan. (2015, November 18). *Kaplan Test Prep and MONEY survey: To pay for college, parents will take staycations, trim retirement savings and defer home improvements, among other financial sacrifices*. Kaplan [Blog]. **https://www.kaptest.com/blog/press/2015/11/18/kaplan-test-prep-and-money-survey-to-pay-for-college-parents-will-take-staycations-trim-retirement-savings-and-defer-home-improvements-among-other-financial-sacrifices**

Kaput, J. J. (1995). Long-term algebra reform: Democratizing access to big ideas. In C. B. Lacampagne, W. Blair, & J. Kaput (Eds.), *The Algebra Initiative Colloquium*, vol. 1 (pp. 33–52). U.S. Department of Education, Office of Educational Research and Improvement. **https://files.eric.ed.gov/fulltext/ED385436.pdf**

Kendeou, P. (2015). A general inference skill. In E. J. O'Brien, A. E. Cook, & R. F. Lorch Jr. (Eds.), *Inferences during reading* (pp. 160–81). Cambridge University Press.

Kieran, C., Pang, J., Schifter, D., & Ng, S. F. (2016). *Early algebra: Research into its nature, its learning, its teaching*. SpringerOpen. **https://library.oapen.org/handle/20.500.12657/27822**

Kim, K., Bae, J., Nho, M., & Lee, C. H. (2011). How do experts and novices differ? Relation versus attribute and thinking versus feeling in language use. *Psychology of Aesthetics, Creativity, and the Arts*, *5*(4), 379–88. **https://doi.org/10.1037/a0024748**

Kjelvik, M. K., & Schultheis, E. H. (2019). Getting messy with authentic data: Exploring the potential of using data from scientific research to support student data literacy. *CBE—Life Sciences Education*, *18*(2). **https://doi.org/10.1187/cbe.18-02-0023**

Klebs, S., Fishman, R., Nguyen, S., & Hiler, T. (2021, June 29). *One year later: COVID-19s impact on current and future college students*. Third Way. **https://www.thirdway.org/memo/one-year-later-covid-19s-impact-on-current-and-future-college-students**

Kobrin, J. L., Patterson, B. F., Shaw, E. J., Mattern, K. D., & Barbuti, S. M. (2008). *Validity of the SAT for predicting first-year college grade point average*. Research report no. 2008-5, College Board. **https://eric.ed.gov/?id=ED563202**

Kolen, M. J., & Brennan, R. L. (2004). *Test equating, scaling, and linking: Methods and practices* (2nd ed.). Springer.

Kuperberg, G. R., Paczynski, M., & Ditman, T. (2011). Establishing causal coherence across sentences: An ERP study. *Journal of Cognitive Neuroscience*, *23*(5), 1230–46. **https://doi.org/10.1162/jocn.2010.21452**

Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age of acquisition ratings for 30,000 English words. *Behavior Research Methods*, *44*(4), 978–90. **https://doi.org/10.3758/s13428-012-0210-4**

Kwon, N. (2017). How work positions affect the work activity and information behaviour of laboratory scientists in the research lifecycle: Applying activity theory. *Information Research*, *22*(1), paper 744. **https://files.eric.ed.gov/fulltext/EJ1138647.pdf**

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, *104*(2), 211–40. **https://psycnet.apa.org/doi/10.1037/0033-295X.104.2.211**

Lapp, D., Moss, B., Grant, M., & Johnson, K. (2015). *A close look at close reading: Teaching students to analyze complex texts, grades K–5*. ASCD. **https://files.ascd.org/staticfiles/ascd/pdf/siteASCD/publications/books/close-look-at-close-reading-k-5-sample-chapters.pdf**

Larson, A. A., Britt, M. A., & Kurby, C. A. (2009). Improving students' evaluation of informal arguments. *Journal of Experimental Education*, *77*(4), 339–66. **https://doi.org/10.3200/JEXE.77.4.339-366**

Lawley D. N. (1943). A note on Karl Pearson's selection formulæ. *Proceedings of the Royal Society of Edinburgh, Section A: Mathematics*, *62*(1), 28–30. **https://www.cambridge.org/core/journals/proceedings-of-the-royal-society-of-edinburgh-section-a-mathematics/article/abs/iva-note-on-karl-pearsons-selection-formulae/4A0B37E19E4FB5144E32EDE9A1C6EE99**

Lawrence, J. F., Hagen, A. M., Hwang, J. K., Lin, G., & Lervåg, A. (2019). Academic vocabulary and reading comprehension: Exploring the relationships across measures of vocabulary knowledge. *Reading and Writing*, *32*(2), 285–306. **https://doi.org/10.1007/s11145-018-9865-2**

Lee, C. D. (2007). *Culture, literacy, and learning: Taking bloom in the midst of the whirlwind*. Teachers College Press.

Lee, S. W., & Mao, X. (2021). Algebra by the eighth grade: The association between early study of Algebra I and students' academic success. *International Journal of Science and Mathematics Education*, *19*(6), 1271–89. **https://doi.org/10.1007/s10763-020-10116-3**

Long, M. C., Conger, D., & Iatarola, P. (2012). Effects of high school course-taking on secondary and postsecondary success. *American Educational Research Journal*, *49*(2), 285–322. **https://doi.org/10.3102%2F0002831211431952**

Lord, F. (1980). *Applications of item response theory to practical testing problems*. Routledge. **https://doi.org/10.4324/9780203056615**

Ma, J., & Pender, M. (2023). *Education pays 2023: The benefits of higher education for individuals and society*. College Board. **https://research.collegeboard.org/media/pdf/education-pays-2023.pdf**

Marini, J. P., Westrick, P. A., Young, L., Ng, H., & Shaw, E. J. (2023). *Digital SAT pilot predictive validity study: A first look* [Research report]. College Board. **https://research.collegeboard.org/media/pdf/Digital%20SAT%28R%29%20Pilot%20Predictive%20Validity%20Study%20-%20A%20First%20Look.pdf**

Marini, J. P., Westrick, P. A., Young, L., & Shaw, E. J. (2022). *Digital SAT score relationships with other educational measures: Early convergent validity evidence* [Research report]. College Board. **https://research.collegeboard.org/media/pdf/Digital_SAT_Score_Relationships_with_Other_Educational_Measures.pdf**

Marsh, H. W., Hau, K. T., & Grayson, D. (2005). Goodness of fit in structural equation models. In A. Maydeu-Olivares & J. J. McArdle (Eds.), *Contemporary psychometrics: A festschrift for Roderick P. McDonald* (pp. 275–340). Lawrence Erlbaum Associates.

Masrai, A. (2019). Vocabulary and reading comprehension revisited: Evidence for high-, mid-, and low-frequency vocabulary knowledge. *SAGE Open*, *9*(2), 1–13. **https://doi.org/10.1177/2158244019845182**

Meng, X. (2021). Enhancing (publications on) data quality: Deeper data minding and fuller data confession. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *184*(4), 1161–75. **https://doi.org/10.1111/rssa.12762**

Merriam-Webster. (n.d.). Geometry. In **Merriam-Webster.com** *dictionary*. Retrieved November 17, 2021. **https://www.merriam-webster.com/dictionary/geometry**

Mislevy, R. J., & Riconsente, M. M. (2005). *Evidence-centered assessment design: Layers, structures, and terminology*. PADI Technical Report 9, SRI International. **https://padi.sri.com/downloads/TR9_ECD.pdf**

Morgan, L. A. (2018, July 1). What careers require geometry. *Chron*. **https://work.chron.com/careers-require-geometry-10361.html**

Moses, R. P., & Cobb Jr., C. E. (2001). *Radical equations: Civil rights from Mississippi to the Algebra Project*. Beacon.

Mullis, I. V. S., Martin, M. O., Beaton, A. E., Gonzalez, E. J., Kelly, D. L, & Smith, T. A. (1998). *Mathematics and science achievement in the final year of secondary school: IEA's Third International Mathematics and Science Study (TIMSS)*. TIMSS International Study Center, Boston College. **https://files.eric.ed.gov/fulltext/ED414207.pdf**

Nagy, W. E., Anderson, R. C., & Herman, P. A. (1987). Learning word meanings from context during normal reading. *American Educational Research Journal*, *24*(2), 237–70. **https://doi.org/10.2307/1162893**

Nation, K. (2009). Reading comprehension and vocabulary: What's the connection? In R. K. Wagner, C. Schatschneider, & C. Phythian-Sence (Eds.), *Beyond decoding: The behavioral and biological foundations of reading comprehension* (pp. 176–94). Guilford.

National Academies of Sciences, Engineering, & Medicine. (2018). *Envisioning the data science discipline: The undergraduate perspective, interim report*. National Academies Press. **https://doi.org/10.17226/24886**

National Assessment Governing Board. (2013). *Grade 12 sample mathematics questions*. The Nation's Report Card, National Assessment of Educational Progress, Institute of Education Sciences, U.S. Department of Education. **https://www.nationsreportcard.gov/mathematics/sample-questions/?grade=12**

National Assessment Governing Board. (2017). *Writing framework for the 2017 National Assessment of Educational Progress*. National Assessment Governing Board. **https://www.nagb.gov/content/dam/nagb/en/documents/publications/frameworks/writing/2017-writing-framework.pdf**

National Assessment Governing Board. (2019). *Reading framework for the 2019 National Assessment of Educational Progress*. National Assessment Governing Board. **https://www.nagb.gov/content/dam/nagb/en/documents/publications/frameworks/reading/2019-reading-framework.pdf**

National Center for Education Statistics. (n.d.). *NAEP report card: Reading. National achievement-level results, grade 12, 2019*. The Nation's Report Card, National Assessment of Educational Progress, Institute of Education Sciences, U.S. Department of Education. **https://www.nationsreportcard.gov/reading/nation/achievement/?grade=12**

National Council for the Social Studies. (2013). *The College, Career, and Civic Life (C3) Framework for social studies state standards: Guidance for enhancing the rigor of K–12 civics, economics, geography, and history*. National Council for the Social Studies. **https://www.socialstudies.org/system/files/2022/c3-framework-for-social-studies-rev0617.2.pdf**

National Council of Teachers of Mathematics. (2018). *Catalyzing change in high school mathematics: Initiating critical conversations*. National Council of Teachers of Mathematics. **https://www.nctm.org/uploadedFiles/Standards_and_Positions/executive%20summary.pdf**

National Governors Association (NGA) Center for Best Practices, & Council of Chief State School Officers (CCSSO). (2010a). *Common Core State Standards for English Language Arts and Literacy in History/Social Studies, Science, and Technical Subjects*. NGA Center for Best Practices. **http://www.corestandards.org/ELA-Literacy**

National Governors Association (NGA) Center for Best Practices, & Council of Chief State School Officers (CCSSO). (2010b). *Common Core State Standards for Mathematics*. NGA Center for Best Practices. **http://www.corestandards.org/wp-content/uploads/Math_Standards1.pdf**

National Mathematics Advisory Panel. (2008). *Foundations for success: The final report of the National Mathematics Advisory Panel*. U.S. Department of Education. **https://files.eric.ed.gov/fulltext/ED500486.pdf**

National Science Foundation. (1976). *Reviews of data on science resources, no. 27: Education and work activities of federal scientific and technical personnel, January 1974*. NSF-76-308. National Institute of Education, U.S. Department of Health, Education, and Welfare. **https://files.eric.ed.gov/fulltext/ED137062.pdf**

National Student Clearinghouse Research Center. (2022). *Overview: Fall 2021 enrollment estimates*. National Student Clearinghouse. **https://nscresearchcenter.org/wp-content/uploads/CTEE_Report_Fall_2021.pdf**

National Student Clearinghouse Research Center. (2023). *Current term enrollment estimates, expanded edition: Fall 2022 enrollment overview*. National Student Clearinghouse. **https://public.tableau.com/app/profile/researchcenter/viz/CTEE_Fall2022_Report/CTEEFalldashboard**

National Student Clearinghouse Research Center. (2024). *Current term enrollment estimates, expanded edition: Fall 2023 enrollment overview*. National Student Clearinghouse. **https://public.tableau.com/app/profile/researchcenter/viz/CTEEFall2023dashboard/CTEEFall2023**

Nelms, A. A., & Segura-Totten, M. (2019). Expert-novice comparison reveals pedagogical implications for students' analysis of primary literature. *CBE—Life Sciences Education 18*(4). **https://doi.org/10.1187/cbe.18-05-0077**

Nelson, J., Perfetti, C., Liben, D, & Liben, M. (2012). *Measures of text difficulty: Testing their predictive value for grade levels and student performance* [Report to the Gates Foundation]. Student Achievement Partners. **https://achievethecore.org/page/1196/measures-of-text-difficulty-testing-their-predictive-value-for-grade-levels-and-student-performance**

Next Generation Science Standards (NGSS) Lead States. (2013). *Next Generation Science Standards: For states, by states*. National Academies Press. **https://nap.nationalacademies.org/catalog/18290/next-generation-science-standards-for-states-by-states**

Oakhill, J. V. (1984). Inferential and memory skills in children's comprehension of stories. *British Journal of Educational Psychology*, *54*(1), 31–39. **https://doi.org/10.1111/j.2044-8279.1984.tb00842.x**

Oakhill, J., & Cain, K. (2012). The precursors of reading ability in young readers: Evidence from a four-year longitudinal study. *Scientific Studies of Reading*, *16*(2), 91–121. **https://doi.org/10.1080/10888438.2010.529219**

O'Connor, R. E. (2014). *Teaching word recognition: Effective strategies for students with learning difficulties* (2nd ed.). Guilford.

O'Neill, K. S. (2018). Applying the Pareto Principle to the analysis of students' errors in grammar, mechanics and style. *Research in Higher Education Journal*, *34*, 1–12. **https://eric.ed.gov/?id=EJ1178476**

Osana, H. P., Lacroix, G. L., Tucker, B. J., Idan, E., & Jabbour, G. (2007). The impact of print exposure quality and inference construction on syllogistic reasoning. *Journal of Educational Psychology*, *99*(4), 888–902. **https://doi.org/10.1037/0022-0663.99.4.888**

Papashane, M., & Hlalele, D. (2014). Academic literacy: A critical cognitive catalyst towards the creation of sustainable learning ecologies in higher education. *Mediterranean Journal of Social Sciences*, *5*(10), 661–71. **http://dx.doi.org/10.5901/mjss.2014.v5n10p661**

Pearson, P. D., Hiebert, E. H., & Kamil, M. L. (2007). Vocabulary assessment: What we know and what we need to learn. *Reading Research Quarterly*, *42*(2), 282–96. **https://doi.org/10.1598/RRQ.42.2.4**

Peffer, M. E., & Ramezani, N. (2019). Assessing epistemological beliefs of experts and novices via practices in authentic science inquiry. *International Journal of STEM Education*, *6*, Article 3. **https://doi.org/10.1186/s40594-018-0157-9**

Pence, H. E. (2014). What is Big Data and why is it important? *Journal of Educational Technology Systems*, *43*(2), 159–71. **https://doi.org/10.2190%2FET.43.2.d**

Phillips, R. J. (1997). Can juniors read graphs? A review and analysis of some computer-based activities. *Journal of Information Technology for Teacher Education*, *6*(1), 49–58. **https://doi.org/10.1080/14759399700200005**.

Picone, C., Rhode, J., Hyatt, L., & Parshall, T. (2007). Assessing gains in undergraduate students' abilities to analyze graphical data. *Teaching Issues and Experiments in Ecology 5*, 1–54. **https://www.esa.org/tiee/vol/v5/research/picone/pdf/Picone_etal2007.pdf**

Preto-Bay, A. M. (2004). The socio-cultural dimension of academic literacy development and the explicit teaching of genres as community heuristics. *Reading Matrix*, *4*(3), 86–117.

Probosari, R. M., Widyastuti, F., Sajidan, Suranto, & Prayitno, B. A. (2019). Students' argument style through scientific reading-based inquiry: Improving argumentation skill in higher education. Second International Conference on Science, Mathematics, Environment, and Education, Surakarta, Indonesia, July 26–28, 2019. AIP Conference Proceedings 2194, 020088, December 18, 2019. **https://doi.org/10.1063/1.5139820**

Quinn, J. M., Wagner, R. K., Petscher, Y., & Lopez, D. (2015). Developmental relations between vocabulary knowledge and reading comprehension: A latent change score modeling study. *Child Development*, *86*(1), 159–75. **https://doi.org/10.1111/cdev.12292**

Rakes, C. R., Valentine, J. C., McGatha, M. B., & Ronau, R. N. (2010). Methods of instructional improvement in algebra: A systematic review and meta-analysis. *Review of Educational Research*, *80*(3), 372–400. **https://doi.org/10.3102/0034654310374880**

Renken, M. D., & Nuñez, N. (2009). Evidence for improved conclusion accuracy after reading about rather than conducting a belief-inconsistent simple physics experiment. *Applied Cognitive Psychology*, *24*(6), 792–811. **https://doi.org/10.1002/acp.1587**

Rice, M. (2023, July 11). *22 big data examples and applications.* Built In (blog). **https://builtin.com/big-data/big-data-examples-applications**

Rose, D. H., Meyer, A., Strangman, N., & Rappolt, G. (2002). *Teaching every student in the digital age: Universal Design for Learning.* Association for Supervision and Curriculum Development.

Rose, H., & Betts, J. R. (2001). *Math matters: The links between high school curriculum, college graduation, and earnings.* Public Policy Institute of California. **https://www.ppic.org/wp-content/uploads/content/pubs/report/R_701JBR.pdf**

Rose, H., & Betts, J. R. (2004). The effect of high school courses on earnings. *Review of Economics and Statistics*, *86*(2), 497–513. **https://doi.org/10.1162/003465304323031076**

Rosenfeld, S. (2019). *Democracy and truth: A short history.* University of Pennsylvania Press.

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph Supplement*, *34*(4, pt. 2), 100. **https://psycnet.apa.org/record/1972-04809-001**

Schleicher, A. (2019). *Programme for International Student Assessment (PISA) 2018: Insights and interpretations.* Organisation for Economic Co-operation and Development (OECD). **https://www.oecd.org/pisa/PISA%202018%20Insights%20and%20Interpretations%20FINAL%20PDF.pdf**

Schmidt, F.L., & Hunter, J.C. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, *124*, 262–74. **https://doi.org/10.1037/0033-2909.124.2.262**

Segers, E., & Verhoeven, L. (2016). How logical reasoning mediates the relation between lexical quality and reading comprehension. *Reading and Writing*, *29*, 577–90. **https://doi.org/10.1007/s11145-015-9613-9**

Shanahan, T. (n.d.). Common Core: Close reading. *Instructor*. Retrieved November 23, 2018. **https://www.shanahanonliteracy.com/publications/instructor-magazine-common-core-close-reading**

Shaw, E. J., Marini, J. P., Beard, J., Shmueli, D., Young, L., & Ng, H. (2016). *The redesigned SAT pilot predictive validity study: A first look* [Research report]. College Board. **https://satsuite.collegeboard.org/media/pdf/redesigned-sat-pilot-predictive-validity-study-first-look.pdf**

Silva, M., & Cain, K. (2015). The relations between lower and higher level comprehension skills and their role in prediction of early reading comprehension. *Journal of Educational Psychology*, *107*(2), 321–31. **https://doi.org/10.1037/a0037769**

Sireci, S. G. (2013). Agreeing on validity arguments. *Journal of Educational Measurement*, *50*, 99–104. **https://doi.org/10.1111/jedm.12005**

Snow, C. E. (2010). Academic language and the challenge of reading for learning about science. *Science 328*(5977), 450–452. **http://dx.doi.org/10.1126/science.1182597**

Snyder, A. J. (2022, October 11). More than 3 in 4 Americans believe college is difficult to afford. *Morning Consult*. **https://morningconsult.com/2022/10/11/college-affordability-student-loans-survey-data**

Stahl, S. A. (2003). Vocabulary and readability: How knowing word meanings affects comprehension. *Topics in Language Disorders*, *23*(3), 241–47. **https://doi.org/10.1097/00011363-200307000-00009**

Stein, M. K., Kaufman, J. H., Sherman, M., & Hillen, A. F. (2011). Algebra: A challenge at the crossroads of policy and practice. *Review of Educational Research*, *81*(4), 453–92. **https://doi.org/10.3102%2F0034654311423025**

Stigler, J. W., Lee, S., & Stevenson, H. W. (1990). *Mathematical knowledge of Japanese, Chinese, and American elementary school children*. National Council of Teachers of Mathematics. **https://eric.ed.gov/?id=ED325398**

Stylianides, G. J., Stylianides, A. J., & Weber, K. (2017). Research on the teaching and learning of proof: Taking stock and moving forward. In J. Cai (Ed.), *Compendium for research in mathematics education* (pp. 237–266). National Council of Teachers of Mathematics.

Tairab, H. H., & Khalaf Al-Naqbi, A. K. (2004). How do secondary school science students interpret and construct scientific graphs? *Journal of Biological Education*, *38*(3), 127–32. **https://doi.org/10.1080/00219266.2004.9655920**

Tenopir, C., King, D. W., & Bush, A. (2004). Medical faculty's use of print and electronic journals: Changes over time and in comparison with scientists. *Journal of the Medical Library Association*, *92*(2), 233–41. **https://www.ncbi.nlm.nih.gov/pmc/articles/PMC385305**

Thissen, D. & Wainer, H. (2001). *Test scoring*. Routledge. **https://doi.org/10.4324/9781410604729**

Thompson, S. J., Johnstone, C. J., & Thurlow, M. L. (2002). *Universal design applied to large scale assessments*. Synthesis report 44. University of Minnesota, National Center on Educational Outcomes. **https://nceo.umn.edu/docs/onlinepubs/synth44.pdf**

Trusty, J., & Niles, S. G. (2004). Realized potential or lost talent: High school variables and bachelor's degree completion. *The Career Development Quarterly*, *53*(1), 2–15. **https://doi.org/10.1002/j.2161-0045.2004.tb00651.x**

U.S. Bureau of Labor Statistics. (2021). *Top ten fastest growing occupations, excluding pandemic recovery* (Graph). U.S. Bureau of Labor Statistics. Retrieved April 2021, from **https://www.bls.gov/emp/graphics/fastest-growing-occupations.htm**

van den Broek, P., & Helder, A. (2017). Cognitive processes in discourse comprehension: Passive processes, reader-initiated processes, and evolving mental representations. *Discourse Processes*, *54*(5–6), 360–72. **https://doi.org/10.1080/0163853X.2017.1306677** von der Mühlen, S., Richter, T., Schmid, S., & Berthold, K. (2019). How to improve argumentation comprehension in university students: Experimental test of a training approach. *Instructional Science*, *47*, 215–37. **https://doi.org/10.1007/s11251-018-9471-3**

Weedmark, D. (2018, June 29). A list of careers that use algebra. *Chron*. **https://work.chron.com/list-careers-use-algebra-14592.html**

Westrick, P. A., Marini, J. P., Young, L., Ng, H., & Shaw, E. J. (2023). *Digital SAT pilot predictive validity study: A comprehensive analysis of first-year college outcomes* [Research report]. College Board. **https://research.collegeboard.org/media/pdf/Digital%20SAT%20Pilot%20Predictive%20Validity%20Study-A%20Comprehensive%20Analysis%20of%20First-Year%20College%20Outcomes.pdf**

Westrick, P. A., Marini, J. P., Young, L., Ng, H., Shmueli, D., & Shaw, E. J. (2019). *Validity of the SAT for predicting first-year grades and retention to the second year* [Research report]. College Board. **https://satsuite.collegeboard.org/media/pdf/national-sat-validity-study.pdf**

Wexler, N. (2019). *The knowledge gap: The hidden cause of America's broken education system—and how to fix it.* Avery.

Whipple, G. M. (Ed.) (1925). *The twenty-fourth yearbook of the National Society for the Study of Education, part 1: Report of the National Committee on Reading*. Public School Publishing.

Wilkerson, T. (2020, November). *Statistics education: An imperative for our future.* National Council of Teachers of Mathematics President's Messages: Trena Wilkerson. NCTM. **https://www.nctm.org/News-and-Calendar/Messages-from-the-President/President_s-Messages_-Trena-Wilkerson/?ps=20&cp=2**

Williams, J. M., & Bizup, J. (2017). *Style: Lessons in clarity and grace* (12th ed.). Pearson.

Willingham, D. T. (2019). *How to teach critical thinking*. Education: Future Frontiers, Occasional Paper Series, New South Wales (Australia) Department of Education. **http://www.danielwillingham.com/uploads/5/0/0/7/5007325/willingham_2019_nsw_critical_thinking2.pdf**

Wineburg, S. S. (1991). Historical problem solving: A study of the cognitive processes used in the evaluation of documentary and pictorial evidence. *Journal of Educational Psychology*, *83*(1), 73–87. **https://doi.org/10.1037/0022-0663.83.1.73**

Wolf, M. (2018). *Reader, come home: The reading brain in a digital world*. Harper.

Zohar, A., & Nemet, F. (2002). Fostering students' knowledge and argumentation skills through dilemmas in human genetics. *Journal of Research in Science Teaching*, *39*(1), 35–62. **https://doi.org/10.1002/tea.10008**

DIGITAL SAT SUITE OF ASSESSMENTS

# Technical Manual

## Appendices

Appendix A 204

Appendix B 226

# Appendix A

Table A-1 summarizes the skill/knowledge testing points assessed as part of the digital SAT Suite's Reading and Writing section. Except for "Text Complexity," which describes a characteristic of the section as a whole, the boldface entries represent the section's content domains and are accompanied by their respective question types and subtypes.

**TABLE A-1** DIGITAL SAT SUITE READING AND WRITING SECTION: SKILL/KNOWLEDGE TESTING POINTS

| Content Dimension | Description |
|---|---|
| **Text Complexity** | The passages (and pairs of passages) on the Reading and Writing section represent a range of text complexities from grades 6–8 through grades 12–14. (Grades 12–14 passages are excluded from appearing on PSAT 8/9.) |
| **Information and Ideas** | Students will use comprehension, analysis, and reasoning skills and knowledge as well as what is stated and implied in texts (including in any accompanying informational graphics) to locate, interpret, evaluate, and integrate information and ideas. |
| **Central Ideas and Details** | Students will determine the central idea of a text and/or interpret the key details supporting that idea. |
| **Command of Evidence** | Students will determine the evidence in a text that best supports a specified claim or point. |
| **Textual** | Students will determine the textual evidence (e.g., a fact, detail, or example from a text) that best supports a specified claim or point. |
| **Quantitative** | Students will determine the quantitative evidence (i.e., data from an informational graphic) that best supports a specified claim or point. |
| **Inferences** | Students will draw reasonable inferences based on explicit and/or implicit information and ideas in a text. |
| **Craft and Structure** | Students will use comprehension, vocabulary, analysis, synthesis, and reasoning skills and knowledge to use and determine the meaning of high-utility academic words and phrases in context, evaluate texts rhetorically, and make supportable connections between multiple topically related texts. |
| **Words in Context** | Students will determine the meaning of a high-utility academic word or phrase in context or use such vocabulary in a contextually appropriate way. |
| **Text Structure and Purpose** | Students will analyze the structure of a text or determine the main rhetorical purpose of a text. |
| **Cross-Text Connections** | Students will draw reasonable connections between two texts on related topics |

| Content Dimension | Description |
|---|---|
| **Expression of Ideas** | Students will use revision skills and knowledge to improve the effectiveness of written expression in accordance with specified rhetorical goals. |
| **Rhetorical Synthesis** | Students will strategically integrate information and ideas on a topic to form an effective sentence achieving a specified rhetorical aim. |
| **Transitions** | Students will determine the most effective transition word or phrase to logically connect information and ideas in a text. |
| **Standard English Conventions** | Students will use editing skills and knowledge to make text conform to core conventions of Standard English sentence structure, usage, and punctuation. |
| **Boundaries** | Students will edit text to ensure that sentences are conventionally complete. |
| **Between Sentences** | Students will use contextually appropriate punctuation to properly mark the end of a sentence. |
| **Within Sentences** | Students will coordinate clauses within a sentence or elements of a series using appropriate punctuation and, in some cases, a conjunction or conjunctive adverb; incorporate supplementary information (e.g., appositives, parentheticals) using appropriate punctuation; and recognize circumstances in which no punctuation is needed to set off sentence elements. |
| **Form, Structure, and Sense** | Students will edit text to conform to conventional usage. |
| **Subject-Verb Agreement** | Students will ensure agreement in number between a subject and a verb. |
| **Pronoun-Antecedent Agreement** | Students will ensure agreement in number between a pronoun and its antecedent. |
| **Verb Finiteness** | Students will use verbs and verbals (i.e., gerunds, participles, infinitives) in contextually appropriate ways. |
| **Verb Tense and Aspect** | Students will use contextually appropriate tenses and aspects of verbs. |
| **Subject-Modifier Placement** | Students will place modifying elements in sentences (e.g., participles) in contextually appropriate ways. |
| **Genitives and Plurals** | Students will make contextually appropriate choices among singular, plural, singular possessive, and plural possessive nouns and pronouns and among possessive determiners (its, their, your), contractions (it's, they're, you're), and adverbs (there). |

Table A-2 summarizes the (operational) distribution of Reading and Writing section questions by content domain. Note that pretest questions aren't considered in the presented ranges.

**TABLE A-2 DIGITAL SAT SUITE READING AND WRITING SECTION: OPERATIONAL QUESTION DISTRIBUTION**

| Content Domain | Domain Description (Claim) | Skill/Knowledge Testing Points | Operational Question Distribution |
|---|---|---|---|
| **Information and Ideas** | Students will use comprehension, analysis, and reasoning skills and knowledge as well as what is stated and implied in texts (including in any accompanying informational graphics) to locate, interpret, evaluate, and integrate information and ideas. | Central Ideas and Details<br>Command of Evidence<br>Textual<br>▪ Quantitative<br>▪ Inferences | ≈26%<br>12–14 questions |
| **Craft and Structure** | Students will use comprehension, vocabulary, analysis, synthesis, and reasoning skills and knowledge to use and determine the meaning of high-utility academic words and phrases in context, evaluate texts rhetorically, and make supportable connections between multiple topically related texts. | Words in Context<br>Text Structure and Purpose<br>Cross-Text Connections | ≈28%<br>13–15 questions |
| **Expression of Ideas** | Students will use revision skills and knowledge to improve the effectiveness of written expression in accordance with specified rhetorical goals. | Rhetorical Synthesis<br>Transitions | ≈20%<br>8–12 questions |
| **Standard English Conventions** | Students will use editing skills and knowledge to make text conform to core conventions of Standard English sentence structure, usage, and punctuation. | Boundaries<br>Form, Structure, and Sense | ≈26%<br>11–15 questions |

Table A-3 (Algebra), Table A-4 (Advanced Math), Table A-5 (Problem-Solving and Data Analysis), and Table A-6 (Geometry and Trigonometry / Geometry) list the content dimensions assessed at each program level: digital SAT, digital PSAT/NMSQT and PSAT 10, and digital PSAT 8/9.

**TABLE A-3** DIGITAL SAT SUITE MATH SECTION: SKILL/KNOWLEDGE TESTING POINTS—ALGEBRA CONTENT DOMAIN

| Content Dimension | SAT Description | PSAT/NMSQT and PSAT 10 Description | PSAT 8/9 Description |
|---|---|---|---|
| **Linear Equations in One Variable** | Create and use linear equations in one variable to solve problems in a variety of contexts. | Create and use linear equations in one variable to solve problems in a variety of contexts. | Create and use linear equations in one variable to solve problems in a variety of contexts. |
| | Identify or create a linear equation in one variable that represents a context. | Identify or create a linear equation in one variable that represents a context. | Identify or create a linear equation in one variable that represents a context. |
| | For a linear equation in one variable, interpret a constant, variable, factor, term, or the solution in a context. | For a linear equation in one variable, interpret a constant, variable, factor, term, or the solution in a context. | For a linear equation in one variable, interpret a constant, variable, factor, term, or the solution in a context. |
| | Solve a linear equation in one variable, making strategic use of algebraic structure. | Solve a linear equation in one variable, making strategic use of algebraic structure. | Solve a linear equation in one variable, making strategic use of algebraic structure. |
| | For a linear equation in one variable, determine the conditions under which the equation has no solution, a unique solution, or infinitely many solutions. | For a linear equation in one variable, determine the conditions under which the equation has no solution, a unique solution, or infinitely many solutions. | Fluently solve a linear equation in one variable. |
| | Fluently solve a linear equation in one variable. | Fluently solve a linear equation in one variable. | |

| Content Dimension | SAT Description | PSAT/NMSQT and PSAT 10 Description | PSAT 8/9 Description |
|---|---|---|---|
| **Linear Functions** | Algebraically, a linear function can be defined by a linear expression in one variable or by a linear equation in two variables. In the first case, the variable is the input and the value of the expression is the output. In the second case, one of the variables is designated as the input and determines a unique value of the other variable, which is the output.<br>▪ Create and use linear functions to solve problems in a variety of contexts.<br>▪ Identify or create a linear function to model a relationship between two quantities.<br>▪ For a linear function that represents a context, interpret the meaning of an input/output pair, constant, variable, factor, or term based on the context, including situations where seeing structure provides an advantage.<br>▪ Interpret the graph of a linear function in a context.<br>▪ Make connections between a table, an algebraic representation, or a graph of a linear function not in context.<br>▪ Make connections between a table, an algebraic representation, or a graph of a linear function in context.<br>▪ For a linear function that represents a context, given an input value, find and interpret the output value using the given representation, or given an output value, find and interpret the input value using the given representation, if it exists.<br>▪ Write the rule for a linear function given two input/output pairs or one input/output pair and the rate of change.<br>▪ Evaluate a linear function given an input value, or find the input value for a corresponding output. | Algebraically, a linear function can be defined by a linear expression in one variable or by a linear equation in two variables. In the first case, the variable is the input and the value of the expression is the output. In the second case, one of the variables is designated as the input and determines a unique value of the other variable, which is the output.<br>▪ Create and use linear functions to solve problems in a variety of contexts.<br>▪ Identify or create a linear function to model a relationship between two quantities.<br>▪ For a linear function that represents a context, interpret the meaning of an input/output pair, constant, variable, factor, or term based on the context, including situations where seeing structure provides an advantage.<br>▪ Interpret the graph of a linear function in a context.<br>▪ Make connections between a table, an algebraic representation, or a graph of a linear function not in context.<br>▪ Make connections between a table, an algebraic representation, or a graph of a linear function in context.<br>▪ For a linear function that represents a context, given an input value, find and interpret the output value using the given representation, or given an output value, find and interpret the input value using the given representation, if it exists.<br>▪ Write the rule for a linear function given two input/output pairs or one input/output pair and the rate of change.<br>▪ Evaluate a linear function given an input value, or find the input value for a corresponding output. | Algebraically, a linear function can be defined by a linear expression in one variable or by a linear equation in two variables. In the first case, the variable is the input and the value of the expression is the output. In the second case, one of the variables is designated as the input and determines a unique value of the other variable, which is the output.<br>▪ Create and use linear functions to solve problems in a variety of contexts.<br>▪ Identify or create a linear function to model a relationship between two quantities.<br>▪ For a linear function that represents a context, interpret the meaning of an input/output pair, constant, variable, factor, or term based on the context, including situations where seeing structure provides an advantage.<br>▪ Interpret the graph of a linear function in a context.<br>▪ Make connections between a table, an algebraic representation, or a graph of a linear function not in context.<br>▪ Make connections between a table, an algebraic representation, or a graph of a linear function in context.<br>▪ For a linear function that represents a context, given an input value, find and interpret the output value using the given representation, or given an output value, find and interpret the input value using the given representation, if it exists.<br>▪ Write the rule for a linear function given two input/output pairs or one input/output pair and the rate of change.<br>▪ Evaluate a linear function given an input value, or find the input value for a corresponding output. |

| Content Dimension | SAT Description | PSAT/NMSQT and PSAT 10 Description | PSAT 8/9 Description |
|---|---|---|---|
| **Linear Equations in Two Variables** | A linear equation in two variables can be used to represent a constraint or condition on two variable quantities in situations where neither of the variables is regarded as an input or an output. A linear equation can also be used to represent a straight line in the coordinate plane.<br>▪ Create and use a linear equation in two variables to solve problems in a variety of contexts.<br>▪ Identify or create a linear equation in two variables to model a constraint or condition on two quantities.<br>▪ For a linear equation in two variables that represents a context, interpret a solution, constant, variable, factor, or term based on the context, including situations where seeing structure provides an advantage.<br>▪ Interpret the graph of a linear equation in the form $AxBy=C$ in a context.<br>▪ Make connections between:<br>◆ an algebraic representation and a graph of a linear equation in two variables not in context.<br>◆ a table and an algebraic representation or between a table and a graph of a linear equation in two variables not in context.<br>▪ Make connections between a table, an algebraic representation, or a graph of a linear equation in two variables in a context.<br>▪ For a linear equation in two variables that represents a context, given a value of one quantity in the relationship, find a value of the other, if it exists.<br>▪ Write an equation for a line given two points on the line, one point and the slope of the line, or one point and a parallel or perpendicular line. | A linear equation in two variables can be used to represent a constraint or condition on two variable quantities in situations where neither of the variables is regarded as an input or an output. A linear equation can also be used to represent a straight line in the coordinate plane.<br>▪ Create and use a linear equation in two variables to solve problems in a variety of contexts.<br>▪ Identify or create a linear equation in two variables to model a constraint or condition on two quantities.<br>▪ For a linear equation in two variables that represents a context, interpret a solution, constant, variable, factor, or term based on the context, including situations where seeing structure provides an advantage.<br>▪ Interpret the graph of a linear equation in the form $Ax+By=C$ in a context.<br>▪ Make connections between:<br>◆ an algebraic representation and a graph of a linear equation in two variables not in context.<br>◆ a table and an algebraic representation or between a table and a graph of a linear equation in two variables not in context.<br>▪ Make connections between a table, an algebraic representation, or a graph of a linear equation in two variables in a context.<br>▪ For a linear equation in two variables that represents a context, given a value of one quantity in the relationship, find a value of the other, if it exists.<br>▪ Write an equation for a line given two points on the line, one point and the slope of the line, or one point and a parallel or perpendicular line. | A linear equation in two variables can be used to represent a constraint or condition on two variable quantities in situations where neither of the variables is regarded as an input or an output. A linear equation can also be used to represent a straight line in the coordinate plane.<br>▪ Create and use a linear equation in two variables to solve problems in a variety of contexts.<br>▪ Identify or create a linear equation in two variables to model a constraint or condition on two quantities.<br>▪ For a linear equation in two variables that represents a context, interpret a solution, constant, variable, factor, or term based on the context, including situations where seeing structure provides an advantage.<br>▪ Interpret the graph of a linear equation in the form $Ax+By=C$ in a context.<br>▪ Make connections between:<br>◆ an algebraic representation and a graph of a linear equation in two variables not in context.<br>◆ a table and an algebraic representation or between a table and a graph of a linear equation in two variables not in context.<br>▪ Make connections between a table, an algebraic representation, or a graph of a linear equation in two variables in a context.<br>▪ For a linear equation in two variables that represents a context, given a value of one quantity in the relationship, find a value of the other, if it exists.<br>▪ Write an equation for a line given two points on the line, one point and the slope of the line, or one point and a parallel or perpendicular line. |

| Content Dimension | SAT Description | PSAT/NMSQT and PSAT 10 Description | PSAT 8/9 Description |
|---|---|---|---|
| **Systems of Two Linear Equations in Two Variables** | Create and use a system of two linear equations in two variables to solve problems in a variety of contexts.<br>Identify or create a system of linear equations in two variables to model constraints or conditions on two quantities.<br>Solve a system of two linear equations in two variables, making strategic use of algebraic structure.<br>For a system of linear equations in two variables, determine the conditions under which the system has no solution, a unique solution, or infinitely many solutions.<br>Make connections between an algebraic representation and a graph of a system of linear equations in two variables not in context.<br>Make connections between an algebraic representation and a graph of a system of linear equations in two variables in a context.<br>Fluently solve a system of linear equations in two variables. | Create and use a system of two linear equations in two variables to solve problems in a variety of contexts.<br>Identify or create a system of linear equations in two variables to model constraints or conditions on two quantities.<br>Solve a system of two linear equations in two variables, making strategic use of algebraic structure.<br>For a system of linear equations in two variables, determine the conditions under which the system has no solution, a unique solution, or infinitely many solutions.<br>Make connections between an algebraic representation and a graph of a system of linear equations in two variables not in context.<br>Make connections between an algebraic representation and a graph of a system of linear equations in two variables in a context.<br>Fluently solve a system of linear equations in two variables. | Create and use a system of two linear equations in two variables to solve problems in a variety of contexts.<br>Identify or create a system of linear equations in two variables to model constraints or conditions on two quantities.<br>Solve a system of two linear equations in two variables, making strategic use of algebraic structure.<br>Make connections between an algebraic representation and a graph of a system of linear equations in two variables not in context.<br>Make connections between an algebraic representation and a graph of a system of linear equations in two variables in a context.<br>Fluently solve a system of linear equations in two variables. |
| **Linear Inequalities in One or Two Variables** | Create and use linear inequalities in one or two variables to solve problems in a variety of contexts.<br>Identify or create linear inequalities in one or two variables to model constraints or conditions on two quantities.<br>For linear inequalities in one or two variables, interpret a constant, variable, factor, term, or solution, including situations where seeing structure provides an advantage.<br>Given a linear inequality or system of linear inequalities, interpret a point in the xy-plane in terms of the solution set.<br>Make connections between tabular, algebraic, and graphical representations of linear inequalities in one or two variables by deriving one from the other. | Create and use linear inequalities in one or two variables to solve problems in a variety of contexts.<br>Identify or create linear inequalities in one or two variables to model constraints or conditions on two quantities.<br>For linear inequalities in one or two variables, interpret a constant, variable, factor, term, or solution, including situations where seeing structure provides an advantage.<br>Given a linear inequality or system of linear inequalities, interpret a point in the xy-plane in terms of the solution set.<br>Make connections between tabular, algebraic, and graphical representations of linear inequalities in one or two variables by deriving one from the other. | Create and use linear inequalities in one or two variables to solve problems in a variety of contexts.<br>Identify or create linear inequalities in one or two variables to model constraints or conditions on two quantities.<br>For linear inequalities in one or two variables, interpret a constant, variable, factor, term, or solution, including situations where seeing structure provides an advantage.<br>Given a linear inequality or system of linear inequalities, interpret a point in the xy-plane in terms of the solution set. |

**TABLE A-4** DIGITAL SAT SUITE MATH SECTION: SKILL/KNOWLEDGE TESTING POINTS—ADVANCED MATH CONTENT DOMAIN

| Content Dimension | SAT Description | PSAT/NMSQT and PSAT 10 Description | PSAT 8/9 Description |
|---|---|---|---|
| **Equivalent Expressions** | Make strategic use of algebraic structure and the properties of operations to identify and create equivalent expressions:<br>▪ by factoring polynomials limited to finding a common factor, rewriting binomials that represent a difference of two squares, and rewriting trinomials as the product of two binomials.<br>▪ including rewriting simple rational expressions, rewriting expressions with rational exponents in radical form, and factoring polynomials not included in the preceding bullet.<br>Fluently add, subtract, and multiply polynomials. | Make strategic use of algebraic structure and the properties of operations to identify and create equivalent expressions by factoring polynomials limited to finding a common factor, rewriting binomials that represent a difference of two squares, and rewriting trinomials as the product of two binomials.<br>Fluently add, subtract, and multiply polynomials. | Make strategic use of algebraic structure and the properties of operations to identify and create equivalent expressions by factoring polynomials limited to finding a common factor, rewriting binomials that represent a difference of two squares, and rewriting trinomials as the product of two binomials.<br>Fluently add, subtract, and multiply polynomials. |

| Content Dimension | SAT Description | PSAT/NMSQT and PSAT 10 Description | PSAT 8/9 Description |
|---|---|---|---|
| **Nonlinear Equations in One Variable and Systems of Equations in Two Variables** | Make strategic use of algebraic structure, the properties of operations, and/or reasoning about equality to solve:<br>▪ quadratic equations in one variable presented in a wide variety of forms.<br>▪ linear absolute value equations in one variable or simple rational and radical equations in one variable.<br>▪ polynomial equations in one variable that are written in factored form.<br>Make strategic use of algebraic structure, the properties of operations, and reasoning about equality to solve systems of linear and nonlinear equations in two variables.<br>Determine the conditions under which a quadratic equation has no real solutions, one real solution, or two real solutions.<br>Relate the solutions of a system of a linear and a nonlinear equation in two variables to the graphs of the equations in the system.<br>Given an equation or formula in two or more variables, view it as an equation in a single variable of interest where the other variables are parameters, and solve for the variable of interest.<br>Fluently solve quadratic equations in one variable, written as a quadratic expression in standard form, where using the quadratic formula or completing the square is the most efficient method for solving the equation. | Make strategic use of algebraic structure, the properties of operations, and/or reasoning about equality to solve:<br>▪ quadratic equations in one variable presented in a wide variety of forms.<br>▪ linear absolute value equations in one variable or simple rational and radical equations in one variable.<br>Make strategic use of algebraic structure, the properties of operations, and reasoning about equality to solve systems of linear and nonlinear equations in two variables.<br>Determine the conditions under which a quadratic equation has no real solutions, one real solution, or two real solutions.<br>Relate the solutions of a system of a linear and a nonlinear equation in two variables to the graphs of the equations in the system.<br>Given an equation or formula in two or more variables, view it as an equation in a single variable of interest where the other variables are parameters, and solve for the variable of interest.<br>Fluently solve quadratic equations in one variable, written as a quadratic expression in standard form, where using the quadratic formula or completing the square is the most efficient method for solving the equation. | Make strategic use of algebraic structure, the properties of operations, and/or reasoning about equality to solve quadratic equations in one variable presented in a wide variety of forms.<br>Make strategic use of algebraic structure, the properties of operations, and reasoning about equality to solve systems of linear and nonlinear equations in two variables.<br>Relate the solutions of a system of a linear and a nonlinear equation in two variables to the graphs of the equations in the system.<br>Given an equation or formula in two or more variables, view it as an equation in a single variable of interest where the other variables are parameters, and solve for the variable of interest.<br>Fluently solve quadratic equations in one variable, written as a quadratic expression in standard form, where using the quadratic formula or completing the square is the most efficient method for solving the equation. |

| Content Dimension | SAT Description | PSAT/NMSQT and PSAT 10 Description | PSAT 8/9 Description |
|---|---|---|---|
| **Nonlinear Functions** | Create and use quadratic or exponential functions to solve problems in a variety of contexts.<br>Identify or create an appropriate quadratic or exponential function to model a relationship between quantities.<br>For a quadratic or exponential function that represents a context:<br>▪ interpret the meaning of an input/output pair including an intercept or initial value, including situations where seeing structure provides an advantage.<br>▪ interpret the meaning of a constant, variable, factor, or term based on the context, including situations where seeing structure provides an advantage.<br>For a quadratic or exponential function in a context:<br>▪ interpret a point on the graph.<br>▪ interpret parts of the graph (other than a point or intercept).<br>Make connections between a table, an algebraic representation, or a graph of a:<br>▪ quadratic or exponential function that does not involve a transformation, not in context.<br>▪ polynomial function, simple rational function, or quadratic or exponential function that involves a transformation, not in context.<br>Make connections between a table, an algebraic representation, or a graph of a:<br>▪ quadratic or exponential function that does not involve a transformation, in a context.<br>▪ polynomial function, simple rational function, or other nonlinear function in a context, or a quadratic or exponential function that involves a transformation in a context. | Create and use quadratic or exponential functions to solve problems in a variety of contexts.<br>Identify or create an appropriate quadratic or exponential function to model a relationship between quantities.<br>For a quadratic or exponential function that represents a context:<br>▪ interpret the meaning of an input/output pair including an intercept or initial value, including situations where seeing structure provides an advantage.<br>▪ interpret the meaning of a constant, variable, factor, or term based on the context, including situations where seeing structure provides an advantage.<br>For a quadratic or exponential function in a context:<br>▪ interpret a point on the graph.<br>▪ interpret parts of the graph (other than a point or intercept).<br>Make connections between a table, an algebraic representation, or a graph of a:<br>▪ quadratic or exponential function that does not involve a transformation, not in context.<br>▪ polynomial function, simple rational function, or quadratic or exponential function that involves a transformation, not in context.<br>Make connections between a table, an algebraic representation, or a graph of a:<br>▪ quadratic or exponential function that does not involve a transformation, in a context.<br>▪ polynomial function, simple rational function, or other nonlinear function in a context, or a quadratic or exponential function that involves a transformation in a context. | For a quadratic or exponential function that represents a context, interpret the meaning of an input/output pair including an intercept or initial value, including situations where seeing structure provides an advantage.<br>For a quadratic or exponential function in a context, interpret a point on the graph.<br>Make connections between a table, an algebraic representation, or a graph of a quadratic or exponential function that does not involve a transformation, not in context.<br>Make connections between a table, an algebraic representation, or a graph of a quadratic or exponential function that does not involve a transformation, in a context.<br>Use function notation to represent and interpret input/output pairs:<br>▪ Evaluate a nonlinear function given an input value; or, for a quadratic function, find the input value for a corresponding output. |

| Content Dimension | SAT Description | PSAT/NMSQT and PSAT 10 Description | PSAT 8/9 Description |
|---|---|---|---|
| | Determine the most suitable form of the expression representing the output of the function to display key features for:<br>▪ a quadratic function.<br>▪ an exponential function.<br>Understand and use the fact that for the graph of $y=f(x)$, the solutions to $f(x)=0$ correspond to x-intercepts of the graph and $f(0)$ corresponds to the y-intercept of the graph; make connections between the input/output pairs and points on a graph; interpret this information in a context.<br>Use function notation to represent and interpret input/output pairs:<br>▪ Evaluate a nonlinear function given an input value; or, for a quadratic function, find the input value for a corresponding output.<br>▪ For exponential, polynomial, radical, and rational functions, find the input value for a corresponding output. | Determine the most suitable form of the expression representing the output of the function to display key features for:<br>▪ a quadratic function.<br>▪ an exponential function.<br>Use function notation to represent and interpret input/output pairs:<br>▪ Evaluate a nonlinear function given an input value; or, for a quadratic function, find the input value for a corresponding output.<br>▪ For exponential, polynomial, radical, and rational functions, find the input value for a corresponding output. | |

**TABLE A-5** DIGITAL SAT SUITE MATH SECTION: SKILL/KNOWLEDGE TESTING POINTS—PROBLEM-SOLVING AND DATA ANALYSIS CONTENT DOMAIN

| Content Dimension | SAT Description | PSAT/NMSQT and PSAT 10 Description | PSAT 8/9 Description |
|---|---|---|---|
| **Ratios, Rates, Proportional Relationships, and Units** | Questions will require students to solve problems by using a proportional relationship between quantities, calculating or using a ratio or rate, and/or using units, derived units, and unit conversion.<br>▪ Apply proportional relationships, ratios, and rates in a wide variety of contexts. Examples include, but are not limited to, scale drawings and problems in the natural and social sciences.<br>▪ Solve problems involving derived units, including those that arise from products (e.g., kilowatt-hours) and quotients (e.g., population per square kilometer).<br>▪ Solve problems involving:<br>◆ a one-step unit conversion.<br>◆ a multistep or multidimensional unit conversion.<br>▪ Understand and use the fact that when two quantities are in a proportional relationship, if one changes by a scale factor, then the other also changes by the same scale factor. | Questions will require students to solve problems by using a proportional relationship between quantities, calculating or using a ratio or rate, and/or using units, derived units, and unit conversion.<br>▪ Apply proportional relationships, ratios, and rates in a wide variety of contexts. Examples include, but are not limited to, scale drawings and problems in the natural and social sciences.<br>▪ Solve problems involving derived units, including those that arise from products (e.g., kilowatt-hours) and quotients (e.g., population per square kilometer).<br>▪ Solve problems involving:<br>◆ a one-step unit conversion.<br>◆ a multistep or multidimensional unit conversion.<br>▪ Understand and use the fact that when two quantities are in a proportional relationship, if one changes by a scale factor, then the other also changes by the same scale factor. | Questions will require students to solve problems by using a proportional relationship between quantities, calculating or using a ratio or rate, and/or using units, derived units, and unit conversion.<br>▪ Apply proportional relationships, ratios, and rates in a wide variety of contexts. Examples include, but are not limited to, scale drawings and problems in the natural and social sciences.<br>▪ Solve problems involving derived units, including those that arise from products (e.g., kilowatt-hours) and quotients (e.g., population per square kilometer).<br>▪ Solve problems involving:<br>◆ a one-step unit conversion.<br>◆ a multistep or multidimensional unit conversion.<br>▪ Understand and use the fact that when two quantities are in a proportional relationship, if one changes by a scale factor, then the other also changes by the same scale factor. |
| **Percentages** | Use percentages to solve problems in a variety of contexts:<br>▪ including, but not limited to, discounts, interest, taxes, and tips.<br>▪ including those that involve percent increases and decreases for many different quantities.<br>Understand and use the relationship between percent change and growth factor (5% and 1.05, for example); include percentages greater than or equal to 100%. | Use percentages to solve problems in a variety of contexts:<br>▪ including, but not limited to, discounts, interest, taxes, and tips.<br>▪ including those that involve percent increases and decreases for many different quantities.<br>Understand and use the relationship between percent change and growth factor (5% and 1.05, for example); include percentages greater than or equal to 100%. | Use percentages to solve problems in a variety of contexts:<br>▪ including, but not limited to, discounts, interest, taxes, and tips.<br>▪ including those that involve percent increases and decreases for many different quantities.<br>Understand and use the relationship between percent change and growth factor (5% and 1.05, for example); include percentages greater than or equal to 100%. |

| Content Dimension | SAT Description | PSAT/NMSQT and PSAT 10 Description | PSAT 8/9 Description |
|---|---|---|---|
| **One-Variable Data: Distributions and Measures of Center and Spread** | Analyze and interpret numerical data distributions represented with frequency tables, histograms, dot plots, and box plots.<br>For quantitative variables, calculate, compare, and interpret mean, median, and range.<br>Compare distributions using measures of center and spread, including<br>▪ distributions with different means and the same standard deviations.<br>▪ distributions with different standard deviations.<br>Understand and describe the effect of outliers on mean and median. | Analyze and interpret numerical data distributions represented with frequency tables, histograms, dot plots, and box plots.<br>For quantitative variables, calculate, compare, and interpret mean, median, and range.<br>Compare distributions using measures of center and spread, including:<br>▪ distributions with different means and the same standard deviations.<br>▪ distributions with different standard deviations.<br>Understand and describe the effect of outliers on mean and median. | Analyze and interpret numerical data distributions represented with frequency tables, histograms, dot plots, and box plots.<br>For quantitative variables, calculate, compare, and interpret mean, median, and range.<br>Compare distributions using measures of center and spread, including distributions with different means and the same standard deviations.<br>Understand and describe the effect of outliers on mean and median. |
| **Two-Variable Data: Models and Scatterplots** | Analyze and interpret data represented in a scatterplot, but do not make predictions.<br>Analyze and interpret data represented in a scatterplot to make predictions.<br>Fit linear models to data represented in a scatterplot.<br>Fit quadratic and exponential models to data represented in a scatterplot.<br>Given a relationship between two quantities, read and interpret graphs modeling the relationship.<br>Compare linear and exponential growth. | Analyze and interpret data represented in a scatterplot, but do not make predictions.<br>Analyze and interpret data represented in a scatterplot to make predictions.<br>Fit linear models to data represented in a scatterplot.<br>Fit quadratic and exponential models to data represented in a scatterplot.<br>Given a relationship between two quantities, read and interpret graphs modeling the relationship.<br>Compare linear and exponential growth. | Analyze and interpret data represented in a scatterplot, but do not make predictions.<br>Fit linear models to data represented in a scatterplot.<br>Given a relationship between two quantities, read and interpret graphs modeling the relationship. |

| Content Dimension | SAT Description | PSAT/NMSQT and PSAT 10 Description | PSAT 8/9 Description |
|---|---|---|---|
| **Probability and Conditional Probability** | Use one- and two-way tables, area models, and other representations to find relative frequency, probabilities, and conditional probabilities.<br>▪ Calculate, express, or interpret the probability or conditional probability of an event using a data display showing frequencies for a single variable, a two-way table, an area model, or a description of a situation. Infrequently, given a probability, determine an unknown number in a data display showing frequencies for a single variable, a two-way table, or a description of a situation, including using a probability to determine the frequency of an event. | Use one- and two-way tables, area models, and other representations to find relative frequency, probabilities, and conditional probabilities.<br>▪ Calculate, express, or interpret the probability or conditional probability of an event using a data display showing frequencies for a single variable, a two-way table, an area model, or a description of a situation. Infrequently, given a probability, determine an unknown number in a data display showing frequencies for a single variable, a two-way table, or a description of a situation, including using a probability to determine the frequency of an event. | Use one- and two-way tables, area models, and other representations to find relative frequency, probabilities, and conditional probabilities.<br>▪ Calculate, express, or interpret the probability or conditional probability of an event using a data display showing frequencies for a single variable, a two-way table, an area model, or a description of a situation. Infrequently, given a probability, determine an unknown number in a data display showing frequencies for a single variable, a two-way table, or a description of a situation, including using a probability to determine the frequency of an event. |
| **Inference from Sample Statistics and Margin of Error** | Use sample mean and sample proportion to estimate population mean and population proportion.<br>Interpret margin of error. Understand that a larger sample size generally leads to a smaller margin of error. | Use sample mean and sample proportion to estimate population mean and population proportion. | |
| **Evaluating Statistical Claims: Observational Studies and Experiments** | With random samples, identify or describe which population the results can be extended to. Given a description of a study with or without random assignment, determine whether there is evidence for a causal relationship.<br>Understand why random assignment provides evidence for a causal relationship in an experimental study.<br>Understand issues related to sampling methods and why a result can be extended only to the population from which the sample was selected. | | |

**TABLE A-6** DIGITAL SAT SUITE MATH SECTION: SKILL/KNOWLEDGE TESTING POINTS—GEOMETRY AND TRIGONOMETRY / GEOMETRY CONTENT DOMAIN

| Content Dimension | SAT Description | PSAT/NMSQT and PSAT 10 Description | PSAT 8/9 Description |
|---|---|---|---|
| **Area and Volume** | Solve real-world and mathematical problems about the:<br>▪ area or perimeter of a geometric figure or an object that can be modeled by a geometric figure using given information.<br>▪ surface area or volume of a geometric figure or an object that can be modeled by a geometric figure using given information such as length, area, surface area, or volume.<br>Apply knowledge that changing by a scale factor of $k$ changes all lengths by a factor of $k$, changes all areas by a factor of $k^2$, and changes all volumes by a factor of $k^3$.<br>Demonstrate procedural fluency by selecting the correct:<br>▪ area formula and correctly calculating a specified value.<br>▪ surface area or volume formula and correctly calculating a specified value. | Solve real-world and mathematical problems about the:<br>▪ area or perimeter of a geometric figure or an object that can be modeled by a geometric figure using given information.<br>▪ surface area or volume of a geometric figure or an object that can be modeled by a geometric figure using given information such as length, area, surface area, or volume.<br>Apply knowledge that changing by a scale factor of $k$ changes all lengths by a factor of $k$, changes all areas by a factor of $k^2$, and changes all volumes by a factor of $k^3$.<br>Demonstrate procedural fluency by selecting the correct:<br>▪ area formula and correctly calculating a specified value.<br>▪ surface area or volume formula and correctly calculating a specified value. | Solve real-world and mathematical problems about the:<br>▪ area or perimeter of a geometric figure or an object that can be modeled by a geometric figure using given information.<br>▪ surface area or volume of a geometric figure or an object that can be modeled by a geometric figure using given information such as length, area, surface area, or volume.<br>Apply knowledge that changing by a scale factor of $k$ changes all lengths by a factor of $k$, changes all areas by a factor of $k^2$, and changes all volumes by a factor of $k^3$.<br>Demonstrate procedural fluency by selecting the correct:<br>▪ area formula and correctly calculating a specified value.<br>▪ surface area or volume formula and correctly calculating a specified value. |
| **Lines, Angles, and Triangles** | Use concepts and theorems relating to congruence and similarity of triangles to solve problems.<br>Determine which statements may be required to prove certain relationships or to satisfy a given theorem.<br>Apply knowledge that changing by a scale factor of $k$ changes all lengths by a factor of $k$, but angle measures remain unchanged.<br>Know and directly apply relevant theorems such as the:<br>▪ triangle angle sum theorem.<br>▪ vertical angle theorem and the relationship of angles formed when a transversal cuts parallel lines. | Use concepts and theorems relating to congruence and similarity of triangles to solve problems.<br>Determine which statements may be required to prove certain relationships or to satisfy a given theorem.<br>Apply knowledge that changing by a scale factor of $k$ changes all lengths by a factor of $k$, but angle measures remain unchanged.<br>Know and directly apply relevant theorems such as the:<br>▪ triangle angle sum theorem.<br>▪ vertical angle theorem and the relationship of angles formed when a transversal cuts parallel lines. | Know and directly apply the triangle angle sum theorem. |

| Content Dimension | SAT Description | PSAT/NMSQT and PSAT 10 Description | PSAT 8/9 Description |
|---|---|---|---|
| **Right Triangles and Trigonometry** | Solve problems in a variety of contexts using:<br>▪ the Pythagorean theorem.<br>▪ properties of special right triangles.<br>▪ right triangle trigonometry.<br>Use similarity to calculate values of sine, cosine, and tangent.<br>Solve problems using the relationship between sine and cosine of complementary angles. | Solve problems in a variety of contexts using:<br>▪ the Pythagorean theorem.<br>▪ properties of special right triangles.<br>▪ right triangle trigonometry. | Solve problems in a variety of contexts using the Pythagorean theorem. |
| **Circles** | Use definitions, properties, and theorems relating to circles and parts of circles such as radii, diameters, tangents, angles, arc lengths, and sector areas to solve problems.<br>Solve problems using either radian measure or trigonometric ratios in the unit circle.<br>Create an equation to represent a circle in the xy-plane.<br>Describe how a change to the equation representing a circle affects the graph of the circle in the xy-plane or how a change to the graph of a circle affects the equation that represents the circle.<br>Understand that the ordered pairs that satisfy an equation of the form $(x-h)^2+(y-k)^2=r^2$ form a circle when plotted in the xy-plane.<br>Convert between angle measures in degrees and radians.<br>Complete the square in an equation representing a circle to determine properties of the circle when it is graphed in the xy-plane and use the distance formula in problems related to circles. | | |

Table A-7 (digital SAT), Table A-8 (digital PSAT/NMSQT and PSAT 10), and Table A-9 (digital PSAT 8/9) summarize the (operational) distribution of Math section questions by testing program and then content domain. Note that pretest questions aren't considered in the presented ranges.

**TABLE A-7** DIGITAL SAT MATH SECTION: OPERATIONAL QUESTION DISTRIBUTION BY CONTENT DOMAIN

| Math Section Content Domain | Domain Description (Claim) | Skill/Knowledge Testing Points | Operational Question Distribution |
|---|---|---|---|
| **Algebra** | Students will interpret, create, use, represent, and solve problems using linear representations, and make connections between different representations of linear relationships. | Linear equations in one variable<br>Linear equations in two variables<br>Linear functions<br>Systems of two linear equations in two variables<br>Linear inequalities in one or two variables | ≈35%<br>13–15 questions |
| **Advanced Math** | Students will interpret, rewrite, fluently solve, make strategic use of structure, and create absolute value, quadratic, exponential, polynomial, rational, radical, and other nonlinear equations and make connections between different representations of a nonlinear relationship between two variables. | Equivalent expressions<br>Nonlinear equations in one variable and systems of equations in two variables<br>Nonlinear functions | ≈35%<br>13–15 questions |

| Math Section Content Domain | Domain Description (Claim) | Skill/Knowledge Testing Points | Operational Question Distribution |
| --- | --- | --- | --- |
| **Problem-Solving and Data Analysis** | Using quantitative reasoning, students will fluently solve problems using percentages, proportional relationships, ratios, rates, and units; analyze and interpret distributions of data; use various representations of data to find relative frequency, probabilities, and conditional probabilities; fit models to data and compare linear and exponential growth; and calculate, compare, and interpret mean, median, and range, compare distributions with the same standard deviation, understand basic study design, and interpret margin of error. | Ratios, rates, proportional relationships, and units<br>Percentages<br>One-variable data: distributions and measures of center and spread<br>Two-variable data: models and scatterplots<br>Probability and conditional probability<br>Inference from sample statistics and margin of error<br>Evaluating statistical claims: observational studies and experiments | ≈15%<br>5–7 questions |
| **Geometry and Trigonometry** | Students will solve problems associated with length, area, volume, and scale factors using geometric figures; determine congruence, similarity, and sufficiency using concepts and theorems about vertical angles, triangles, and parallel lines cut by a transversal; solve problems using the Pythagorean theorem, right triangle and unit circle trigonometry, and properties of special right triangles; and use properties and theorems relating to circles to solve problems. | Area and volume<br>Lines, angles, and triangles<br>Right triangles and trigonometry<br>Circles | ≈15%<br>5–7 questions |

**TABLE A-8** DIGITAL PSAT/NMSQT AND PSAT 10 MATH SECTION: OPERATIONAL QUESTION DISTRIBUTION BY CONTENT DOMAIN

| Math Section Content Domain | Domain Description (Claim) | Skill/Knowledge Testing Points | Operational Question Distribution |
|---|---|---|---|
| **Algebra** | Students will interpret, create, use, represent, and solve problems using linear representations and make connections between different representations of linear relationships. | Linear equations in one variable<br>Linear equations in two variables<br>Linear functions<br>Systems of two linear equations in two variables<br>Linear inequalities in one or two variables | ≈35%<br>13–15 questions |
| **Advanced Math** | Students will interpret, rewrite, fluently solve, make strategic use of structure, and create absolute value, quadratic, exponential, polynomial, rational, radical, and other nonlinear equations and make connections between different representations of a nonlinear relationship between two variables. | Equivalent expressions<br>Nonlinear equations in one variable and systems of equations in two variables<br>Nonlinear functions | ≈32.5%<br>12–14 questions |
| **Problem-Solving and Data Analysis** | Using quantitative reasoning, students will fluently solve problems using percentages, proportional relationships, ratios, rates, and units; analyze and interpret distributions of data; use various representations of data to find relative frequency, probabilities, and conditional probabilities; fit models to data and compare linear and exponential growth; and calculate, compare, and interpret mean, median, and range and compare distributions with the same and different standard deviation. | Ratios, rates, proportional relationships, and units<br>Percentages<br>One-variable data: distributions and measures of center and spread<br>Two-variable data: models and scatterplots<br>Probability and conditional probability<br>Inference from sample statistics | ≈20%<br>7–9 questions |

| Math Section Content Domain | Domain Description (Claim) | Skill/Knowledge Testing Points | Operational Question Distribution |
|---|---|---|---|
| **Geometry and Trigonometry** | Students will solve problems associated with length, area, volume, and scale factors using geometric figures; determine congruence, similarity, and sufficiency using concepts and theorems about vertical angles, triangles, and parallel lines cut by a transversal; and solve problems using the Pythagorean theorem and right triangle trigonometry. | Area and volume<br>Lines, angles, and triangles<br>Right triangles and right triangle trigonometry | ≈12.5%<br>4–6 questions |

**TABLE A-9** DIGITAL PSAT 8/9 MATH SECTION: OPERATIONAL QUESTION DISTRIBUTION BY CONTENT DOMAIN.

| Math Section Content Domain | Domain Description (Claim) | Skill/Knowledge Testing Points | Operational Question Distribution |
|---|---|---|---|
| **Algebra** | Students will interpret, create, use, represent, and solve problems using linear representations and make connections between different representations of linear relationships. | Linear equations in one variable<br>Linear equations in two variables<br>Linear functions<br>Systems of two linear equations in two variables<br>Linear inequalities in one or two variables | ≈42.5%<br>16–18 questions |
| **Advanced Math** | Students will rewrite, fluently solve, and make strategic use of structure, absolute value, quadratic, exponential, polynomial, and other nonlinear equations and make connections between different representations of a nonlinear relationship between two variables. | Equivalent expressions<br>Nonlinear equations in one variable and systems of equations in two variables<br>Nonlinear functions | ≈20%<br>7–9 questions |

| Math Section Content Domain | Domain Description (Claim) | Skill/Knowledge Testing Points | Operational Question Distribution |
|---|---|---|---|
| **Problem-Solving and Data Analysis** | Using quantitative reasoning, students will fluently solve problems using percentages, proportional relationships, ratios, rates, and units; analyze and interpret distributions of data; use various representations of data to find relative frequency, probabilities, and conditional probabilities; fit models to data; and calculate, compare, and interpret mean, median, and range. | Ratios, rates, proportional relationships, and units<br>Percentages<br>One-variable data: distributions and measures of center and spread<br>Two-variable data: models and scatterplots<br>Probability and conditional probability | ≈25%<br>9–11 questions |
| **Geometry** | Students will solve problems associated with length, area, volume, and scale factors using geometric figures; apply theorems such as triangle sum; and solve problems using the Pythagorean theorem. | Area and volume<br>Lines, angles, and triangles, including right triangles | ≈12.5%<br>4–6 questions |

Table A-10 summarizes question distribution on the digital SAT Suite's Math section on several dimensions, including by format (multiple-choice or student-produced response), by the presence or absence of context, and by content domain.

**TABLE A-10 DIGITAL SAT SUITE MATH SECTION: OPERATIONAL QUESTION DISTRIBUTION SUMMARY**

| Feature | Digital SAT Suite Testing Program | | |
|---|---|---|---|
| | **SAT** | **PSAT/NMSQT and PSAT 10** | **PSAT 8/9** |
| Operational Questions | 40 | 40 | 40 |
| Questions by Format (%, #) | | | |
| Multiple-Choice (MC) | ≈75% / 28–32 | ≈75% / 28–32 | ≈75% 28–32 |
| Student-Produced Response (SPR) | ≈25% / 8–12 | ≈25% / 8–12 | ≈25% / 8–12 |
| Questions in Context (%, #) | ≈30% / 10–14 | ≈30% / 10–14 | ≈30% / 10–14 |
| Questions by Content Domain | | | |
| Algebra | ≈35% / 13–15 | ≈35% / 13–15 | ≈42.5% / 16–18 |
| Advanced Math | ≈35% / 13–15 | ≈32.5% / 12–14 | ≈20% / 7–9 |
| Problem-Solving and Data Analysis | ≈15% / 5–7 | ≈20% / 7–9 | ≈25% / 9–11 |
| Geometry and Trigonometry (SAT, PSAT/NMSQT, PSAT 10) / Geometry (PSAT 8/9) | ≈15% / 5–7 | ≈12.5% / 4–6 | ≈12.5% / 4–6<br>(Only geometry for PSAT 8/9) |

# Appendix B

**TABLE B-1 INSTITUTIONAL CHARACTERISTICS OF THE STUDY POPULATION FOR SAMPLE RECRUITMENT**

| | Variable | Sample (k=11) | Population (k=788) |
|---|---|---|---|
| **U.S. Region** | Midwest | 9% | 21% |
| | Mid-Atlantic | 9% | 26% |
| | New England | 18% | 11% |
| | South | 27% | 22% |
| | Southwest | 18% | 9% |
| | West | 18% | 10% |
| **Control** | Public | 64% | 40% |
| | Private | 36% | 60% |
| **Admittance rate** | Under 25% | 27% | 6% |
| | 25% to 50% | 9% | 9% |
| | 51% to 75% | 55% | 37% |
| | Over 75% | 9% | 48% |
| **Undergraduate Enrollment** | Small | 0% | 52% |
| | Medium | 0% | 20% |
| | Large | 9% | 14% |
| | Very Large | 91% | 14% |

*Note.* Percentages may not sum to 100 due to rounding. In order to be included in the study population used to develop a representative sampling plan for U.S. four-year institutional participation, institutions had to have at least 250 first-year students, of which at least 75 had SAT scores, at least 15% of first-year students had to have SAT scores, and the institution had to have published admittance rates. Institutions in the U.S. Virgin Islands and Puerto Rico were excluded. Undergraduate enrollment was categorized as follows: small: 4,999 or less; medium: 5,000 to 9,999; large: 10,000 to 19,999; and very large: 20,000 or more.

**TABLE B-2** STUDENT CHARACTERISTICS OF THE CURRENT STUDY SAMPLE, 2020 SAT VALIDITY STUDY SAMPLE, AND THE 2022 COLLEGE-BOUND SENIORS POPULATION

| | Category | Sample (n=1,889) | 2020 Validity Study Sample (n=181,718) | 2022 College-Bound Seniors (n=1,737,678) |
|---|---|---|---|---|
| Gender | Male | 42% | 44% | 48% |
| | Female | 58% | 56% | 51% |
| | Another/Omitted | <1% | 0% | <1% |
| Ethnicity | American Indian/Alaska Native | <1% | <1% | 1% |
| | Asian | 33% | 12% | 10% |
| | Black/African American | 6% | 7% | 12% |
| | Hispanic/Latino | 21% | 16% | 23% |
| | Native Hawaiian/ Other Pacific Islander | <1% | <1% | <1% |
| | White | 33% | 58% | 42% |
| | Two or More Races | 4% | 5% | 4% |
| | No Response | 3% | 2% | 8% |
| Best Language | English Only | 74% | 86% | 71% |
| | English and Another | 24% | 13% | 17% |
| | Another | 2% | 2% | 3% |
| | No Response | <1% | <1% | 9% |
| Highest Parental Education | No High School | 4% | 4% | 6% |
| | High School Diploma | 14% | 16% | 21% |
| | Associate Degree | 4% | 6% | 5% |
| | Bachelor's Degree | 35% | 39% | 28% |
| | Graduate Degree | 39% | 35% | 23% |
| | No Response | 3% | 1% | 17% |