CollegeBoard | ACCUPLACER®

ACCUPLACER®
Suite of
Assessments
Technical Manual

CHARACTERISTICS OF ACCUPLACER

April 2019

Contents

List of Tables

Foreword

The challenge of getting students to and through college is a shared responsibility. Educators, institutions, policymakers, researchers, and test providers must all work together to make sure that the best information is available about what students know and can do so that we can make accurate and fair course placement decisions and offer support services to ensure that all students are on a path to success.

For over 30 years, ACCUPLACER® has been used successfully worldwide, in combination with both academic and non-academic factors, to assess student readiness to succeed in credit-bearing college courses and in workforce training programs. As a computer-adaptive assessment, ACCUPLACER delivers immediate results that support course placement as well as timely interventions with those students whose skills need further development. ACCUPLACER results are used by high schools with dual enrollment programs, community colleges, four-year colleges, and technical schools around the world. The ACCUPLACER Suite of Assessments consists of five tests, Reading; Writing; Arithmetic; Quantitative Reasoning, Algebra, and Statistics; and Advanced Algebra and Functions to place students in college courses; five English as a Second Language (ESL) tests to place college-bound English language learners in appropriate ESL courses; and WritePlacer® tests to evaluate students' direct writing skills.

In 2016, College Board introduced new ACCUPLACER college placement tests that reflect a clearer, research-based focus on what matters most for college readiness and success. The Reading and Writing tests place a greater emphasis on the meaning of words in extended authentic contexts and how word choice shapes meaning, tone, and impact. The redesigned math tests emphasize applied reasoning skills, fluency with mathematical procedures, and conceptual understanding, with a strong acknowledgement that different college major pathways require different levels of mathematics proficiency for success.

This technical manual documents the processes and outcomes of the design and administration of the ACCUPLACER assessments. More importantly, this Manual represents a baseline of evidence supporting the development and psychometric quality of the assessments, including reliability, validity, and fairness of use. This Manual will be supplemented as new evidence is acquired as more and more students nationally and internationally take the assessments. Research-based evidence is the hallmark of College Board's work, and we will continue to evaluate and refine the ACCUPLACER assessments and the models of use to promote student readiness and success as we amass more evidence as part of our ongoing research commitment.

I commend the many members of the College Board team who worked so hard to redesign the ACCUPLACER assessments as well as to develop this technical manual. They were committed to presenting this documentation in an easy-to-read format and to the documentation of clear, concise evidence supporting the purposes and uses of the assessments. Ultimately, those who use the ACCUPLACER program will evaluate whether we achieved these goals. We welcome suggestions for improvement as we continue to supplement this technical manual in the future.

Cynthia Board Schmeiser
Senior Advisor to the CEO
College Board

Contributors

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

Chapter Leads

Luz Bay, Psychometrics

Constance Tsai, Assessment Design and Development

Ashleigh Goldberg, ACCUPLACER Content Management

Writers/Reviewers

Jennifer Bedet, Carly Bonar, Jim Patterson, Thomas Gleiber, Bill Trapp, Thomas Proctor, Siang Chee Chuah, Hui Deng

Technical Manual Working Group

Mark Syp, Chief Editor

Dara Meiner, Group Lead

Luz Bay, Psychometric Lead

Constance Tsai, Assessment Design and Development Lead

Ashleigh Goldberg, ACCUPLACER Content Management Lead

Leadership Reviewers

Antoinette Crockrell, ACCUPLACER Management

Sherral Miller, Assessment Design and Development

Cynthia Schmeiser, Office of the CEO

Special Thanks

James Chandler

Lauren Molin

Preface

Purpose of Manual

The purpose of this technical manual is to provide those interested in using the ACCUPLACER Suite of Assessments with information about the technical qualities of the assessments in the suite. This manual contains information pertaining to the purpose of the assessments—the ACCUPLACER placement tests, ACCUPLACER English as a Second Language (ESL) tests, WritePlacer, and WritePlacer ESL—and the rationale and principles behind their design and development. It describes the content of the assessments; the procedures and processes that are undertaken in the design, development, administration, and scoring of the assessments; the appropriate interpretation of results; the accuracy of the scores from a measurement perspective; and evidence that bears on the validity of interpretations made on the basis of the scores.

College Board believes that it is essential to provide documentation of this nature, in keeping with our organization's commitment to transparency and our desire to adhere to industry best practices and the AERA/APA/NCME standards governing supporting documentation for tests (found in Chapter 7 of the 2014 AERA/APA/NCME *Standards for Educational and Psychological Testing). Maintaining assessments with strong evidence of validity supporting them is an ongoing process. To this end, this manual was conceived as a "living document," and in order to keep the information in the manual as current as possible for users (Standard 7.14), it will be updated as more information becomes available.*

Manual Contents

For ease of reading and understanding, this manual is structured in a manner matching that of "the lifecycle of an ACCUPLACER test." It provides insights about the ACCUPLACER Suite, from the tests' design and through test development, administration, scoring, and the interpretation of those scores for intended uses.

As its name implies, Chapter 1: Overview provides an overview of the ACCUPLACER Suite. It includes a brief history of the assessments in the suite, a discussion of the features of the assessments, their purpose, and several of their intended uses. Chapter 2: Fairness provides an examination of how College Board ensures that ACCUPLACER test content is inclusive, representative, and accessible, consistent with the designs and aims of the tests; the steps taken to ensure tests are administered in a manner that is fair and equitable for all test takers; and the efforts made to support fair and valid interpretation of test scores.

Chapter 3: Test Development Procedures details the process used to develop ACCUPLACER assessments. This includes the establishment of test specifications for each test in the suite; the test development process; the creation, review, and analysis of the questions that make up the tests; the ACCUPLACER computer-adaptive test algorithm, and special accommodations for test takers who require alternative formats.

After detailing the creation of the assessments in the test development portion of the manual, Chapter 4: Testing Requirements describes the procedures used to administer the tests, including test security measures, in a manner that supports fair and valid uses of the tests.

With the assessments having been administered, we then turn our attention toward the scores that are produced. Chapter 5: Interpretation and Application of Results looks at the scoring procedures and analyses used to ensure that placement decisions made using ACCUPLACER scores lead to correct conclusions. Chapter 6: Psychometrics takes this one step further, as it describes procedures necessary to ensure that ACCUPLACER tests are the best assessment possible of that which they are intended to measure.

This manual concludes with Chapter 7: Validity, where the guiding principles of validity are presented. As this manual aims to make apparent, validity considerations permeate every aspect of ACCUPLACER. We have chosen to address validity at this point, as validity evidence takes the results of all of our previous analyses and addresses whether the assessments can be used to determine college readiness, the overall goal of ACCUPLACER. With validity considerations pervading each chapter, emphasis on content-oriented validity is highlighted in this chapter.

Overview

This chapter provides an overview of the ACCUPLACER Suite of Assessments and offers insight into the content of the assessments in the ACCUPLACER Suite. It also presents College Board's motivations for the development, and the guiding principles behind the design, of the latest additions to the ACCUPLACER Suite: ACCUPLACER Placement tests in Reading; Writing; Arithmetic; Quantitative Reasoning, Algebra, and Statistics; and Advanced Algebra and Functions.

Table 1.1 presents an overview of the assessments available in the ACCUPLACER Suite.

Table 1.1: *The ACCUPLACER Suite of Assessments*

ACCUPLACER Placement Tests	Reading
	Writing
	Arithmetic
	Quantitative Reasoning, Algebra, and Statistics
	Advanced Algebra and Functions
ACCUPLACER English as a Second Language (ESL) Tests	Reading Skills
	Sentence Meaning
	Language Use
	Listening
ACCUPLACER Essay Tests	WritePlacer
	WritePlacer ESL

Section 1.1 takes an initial look at the ACCUPLACER Suite. In addition to providing an overview of the features of the assessments in the suite, it discusses the purpose and several of the intended uses of the assessments. The section also includes a brief primer on the concepts of validity, reliability, and fairness, with a specific focus on their relevance to the ACCUPLACER Suite.

Section 1.2 puts the ACCUPLACER Suite into a broader context by providing a brief history of the development of the assessments, including the design of the five placement tests. It also details the benefits the ACCUPLACER Suite brings to students, parents, and K–12 educators. This includes a special focus on the role that student preparation, through challenging high school coursework and test preparation, has on ACCUPLACER outcomes.

Section 1.3 is a description of the content of the ACCUPLACER assessments. We demonstrate how the principles and tenets discussed in the previous sections were applied to create assessments that provide a sound evaluation of student attainment of the skills and knowledge most important to college and career readiness and success.

1.1 Introduction
Brief Description of ACCUPLACER
For over 30 years, ACCUPLACER has been used successfully worldwide, in combination with factors such as high school grade point average (HSGPA), to assess student preparedness for postsecondary education. Tests within the ACCUPLACER Suite of Assessments are delivered over the internet and are used by high schools with dual enrollment programs, community colleges, four-year colleges, and technical schools around the world. Scores from the placement tests are used to inform college placement

decisions and to identify where students are likely to be positioned or likely to succeed within a postsecondary course framework. English as a Second Language (ESL) tests in the ACCUPLACER Suite help place college-bound English language learners in appropriate ESL courses and help propel them toward their academic goals, and the WritePlacer tests provide a measure of test takers' developed ability to write effectively.

Features of ACCUPLACER Assessments

ACCUPLACER Placement Tests

The current ACCUPLACER college placement tests were introduced in September 2016 as part of the College Board Readiness and Success System, a system designed to provide students a platform on which to demonstrate their skills and knowledge, and then to enable schools to provide the support and opportunities that make it easier for students to navigate a path through high school, college, and career. Made up of five evidence-based placement tests (Reading; Writing; Arithmetic; Quantitative Reasoning, Algebra, and Statistics; and Advanced Algebra and Functions), they replace the "classic" ACCUPLACER placement tests (Reading Comprehension, Writing Skills, Arithmetic, Elementary Algebra, and College-Level Math).

The ACCUPLACER placement tests focus comprehensively on the few durable skills that evidence shows matter the most for college and career readiness. These include an emphasis on reasoning alongside a clearer, stronger focus on the skills and knowledge that are the most important for college and career readiness and success. The Reading and Writing Tests also place a greater emphasis on the meaning of words in extended contexts than did their classic counterparts, and on how word choice shapes meaning, tone, and impact. The Math Tests (Arithmetic; Quantitative Reasoning, Algebra, and Statistics; and Advance Algebra and Functions) emphasize applied reasoning skills, fluency with mathematical procedures, conceptual understanding, and a strong connection to multiple mathematics pathways in college curricula. With the exception of the 25-item Writing Test, each test requires students to answer 20 questions. For a more in-depth look at test content and to view specifications, see Chapter 3: Test Development Procedures.

The score reporting for the ACCPLACER placement tests is on a scale ranging from 200 to 300.

English as a Second Language (ESL) Tests

The ACCUPLACER ESL offering is composed of five tests. Four are multiple-choice tests: Reading Skills, Sentence Meaning, Language Use, and Listening. The last, an essay test called WritePlacer ESL, is discussed more in the next subsection.

ACCUPLACER ESL tests are designed for use in placing English language learners (ELLs) into the appropriate ESL courses, where they may receive instruction that supports their path to and success in college. Each multiple-choice ESL test requires students to answer 20 multiple-choice items. They are computer-adaptive and delivered over the internet. Alternate formats are available for students with documented needs for special accommodations. For a more in-depth look at test content and specifications, see Chapter 3: Test Development Procedures.

The score reporting for the ESL tests is on a scale ranging from 20 to 120.

COMPANION Tests

Alternate-format tests called COMPANION™ tests are fixed-length linear forms that are equivalent to the online, computer-adaptive tests in specifications. Each computer-adaptive test has two COMPANION tests. Formats available include braille, audio CD, PDF (paper-and-pencil) and digital (for placement tests).

WritePlacer and WritePlacer ESL

WritePlacer and WritePlacer ESL are the essay tests offered in the ACCUPLACER Suite. Each prompt consists of a short stimulus followed by an assignment that states the writing task. Prompts are designed to be accessible to diverse groups of test takers and

relevant to a wide range of topics and student interests. Test takers' writing samples are electronically scored on the basis of how clearly and effectively they communicate the writers' point of view; the position test takers take on the essay topic is neither rewarded nor penalized. Holistic WritePlacer scores range from 1 to 8, while WritePlacer ESL scores range from 1 to 6.

For both WritePlacer tests, a brief score description is provided along with the holistic score. In addition, if deemed helpful to programs or institutions receiving test scores, feedback may be provided on a range of dimensions considered essential in a well-written essay. See WritePlacer Scoring Rubrics in Appendix A: Scoring Rubrics—WritePlacer and WritePlacer ESL for information on how WritePlacer and WritePlacer ESL essays are scored.

Statement of Purpose:
Intentions and Uses of the ACCUPLACER Suite of Assessments

The primary purpose of the ACCUPLACER Suite of Assessments is to determine the degree to which students are prepared to succeed both in credit-bearing college courses and in workforce training programs. Tests within the ACCUPLACER Suite are designed to collect evidence from student performance in support of a broad claim about what students know and can do. Test scores provide meaningful information about a student's likelihood of success in the resulting placements because the ACCUPLACER placement tests assess content that research shows matters the most for college and career readiness, and the ESL tests assess content and skills that are most relevant for placing test takers into appropriate ESL courses. ACCUPLACER Suite scores should be used in conjunction with other relevant information (e.g., HSGPA and other test scores) to make student course placement decisions. As stated in *The Standards for Educational and Psychological Testing,* Standard 12.10:

> In educational settings, a decision or characterization that will have major impact on a student should take into consideration not just scores from a single test, but other relevant information (AERA, APA, & NCME, 2014, p. 198).

The ACCUPLACER Suite provides data that are used for many purposes by different users. The three key users are higher education, K–12 educators, and students. In keeping with best practices and AERA/APA/NCME standards, the ACCUPLACER Suite's primary intended uses and interpretations for each group of primary users are discussed in the following paragraphs, with a rationale presented for each use. A summary of the evidence and theory bearing on each intended interpretation is presented in Chapter 7: Validity (AERA, APA, & NCME, 2014).

Intended Uses and Interpretations

Proper course placement is critical to a student's success in credit-bearing college courses and in workforce training programs. ACCUPLACER test scores are intended for placing entering college students in the courses that are most appropriate for their level of skill and knowledge. The Reading, Writing, and Math placement tests provide information on a test taker's level of preparedness for postsecondary work that helps postsecondary professionals make informed placement decisions. The ACCUPLACER ESL tests, developed primarily as a response to an expressed need for assessment of the skills and knowledge of English language learners (ELLs), are intended for placing ELLs into the appropriate ESL courses. WritePlacer and WritePlacer ESL are essay tests designed to assess new college students' writing skills; WritePlacer is used to determine whether a test taker has the writing skills required for college-level coursework or would benefit from developmental or corequisite instruction, while WritePlacer ESL is intended to place ELLs into language classes where they may receive the instruction needed to improve their writing skills.

ACCUPLACER may be administered to identify high school students who are eligible for dual enrollment. ACCUPLACER tests would be administered to these high school students late in their junior year or early in their senior year so they have time to work toward meeting requirements for credit-bearing college courses and workforce training programs.

ACCUPLACER Suite scores should not be used as a high school graduation requirement, as an accountability assessment, or as a tool for evaluating teachers, teaching, or educational institutions.

Users should exercise care when attempting to interpret test results for a purpose other than the intended purposes described in this chapter. For further examples of uses of College Board test scores that should be avoided, see the *Guidelines on the Uses of College Board Test Scores and Related Data*, available online (College Board, 2011).[1]

Overview of Fairness, Reliability, and Validity and Their Relevance to ACCUPLACER

As the utility of the ACCUPLACER Suite is so intricately tied to the evidence available on the fairness, reliability, and validity of the associated scores, this manual presents all relevant evidence in these areas to date. This information is interwoven throughout many of the following chapters, and the manual also has dedicated sections on these topics. College Board is committed to adhering to the guidelines and standards outlined in the AERA/APA/NCME standards, which is the preeminent publication in the field of testing for outlining the criteria for test development as well as the evaluation of tests, testing practices, and the interpretation of test scores for their intended uses. As such, this technical manual explains and provides information on fairness, reliability, validity, and other critical topics that test users need to interpret and evaluate the technical quality of the assessments in the ACCUPLACER Suite.

Fairness in testing touches on critical issues such as the equitable treatment of all test takers during the testing process, the absence of measurement bias in test scores, and test taker access to the material that is being assessed, as well as ways of ensuring that the interpretations of individual test scores for intended uses are valid (AERA, APA, & NCME, 2014). Information on the fairness of the ACCUPLACER Suite can be found in essentially all chapters of this manual, in order to highlight its central importance and foundational role in the design, development, administration, scoring, use, and interpretation of the ACCUPLACER Suite and the associated scores. It's also discussed in depth in a chapter on fairness, which provides the reader with a summary of fairness in one location (see Chapter 2: Fairness).

Reliability can be thought of as a prerequisite condition for the interpretation of test scores to be valid for a specific use. Reliability in this manual is further specified when discussed in the Reliability section, but it most generally refers to the "consistency of the scores across instances of the testing procedure" (AERA, APA, & NCME, 2014, p. 33). In this manual, we outline the measures taken to ensure that ACCUPLACER assessment scores are reliable across multiple administrations of the assessment over time and in different contexts, and we document the reliability estimates of the assessment scores (see Section 6.4: Reliability).

When we refer to validity in this manual, it's defined as "the degree to which evidence and theory support the interpretations of test scores for proposed uses of tests" (AERA, APA, & NCME, 2014, p. 11). Note that validity isn't a property of the test itself but rather refers to the interpretation of test scores for a specific use. The concept of validity is woven throughout this manual, as most processes and procedures discussed in this manual play a role in ensuring that the ACCUPLACER assessments have scores that are valid for their intended uses. There is also a distinct chapter on validity that concludes this manual, where we outline the intended interpretations of specific uses of ACCUPLACER Suite scores and provide the related validity evidence available for those uses (see Chapter 7: Validity).

---

1. Available at: **http://media.collegeboard.com/digitalServices/pdf/research/guidelines-on-uses-of-college-board-test-scores-and-data.pdf**

1.2 Brief History of Development

As mentioned previously, the ACCUPLACER Suite is made up of five ACCUPLACER placement tests designed to measure student readiness for credit-bearing college courses, four ESL tests, and two essay tests (see Table 1.1 for a summary).

Background on Assessment

Beginning in 2014, five evidence-based ACCUPLACER college-placement tests were developed to replace existing placement tests. In September 2016, these were introduced to ACCUPLACER users nationwide as part of the current ACCUPLACER Suite, to be offered alongside existing ACCUPLACER tests. The new ACCUPLACER Suite is made up of the new placement tests, existing ACCUPLACER ESL tests, WritePlacer, and WritePlacer ESL.

The initial impetus for the development of the ACCUPLACER ESL tests came from findings from a 1988 nationwide survey conducted by College Board, which revealed a need for an ESL placement test for two-year colleges. In 1993, three initial ESL tests, then called Levels of English Proficiency (LOEP) Tests, were offered as a response to these findings. Each of these, LOEP Sentence Meaning, Language Use, and Reading Skills, were designed to assist with the placement of ELLs into ESL courses. In 2002, LOEP Listening and WritePlacer ESL were added. In 2005, the LOEP tests were renamed the English as a Second Language tests.

Following is a brief timeline:

1985 – College Board introduced the ACCUPLACER Computerized Placement Tests to assist with the placement of entering college students into English and math courses. There were four tests in the initial offering: Reading Comprehension, Sentence Skills, Arithmetic, and Elementary Algebra.

1990 – College-Level Math, designed to assist with the placement of students into upper-level math courses, was introduced.

1993 – Three LOEP tests were introduced to assist with the placement of students into ESL courses.

2002 – LOEP Listening and WritePlacer ESL were added. Also introduced was WritePlacer, the essay test for non-ELLs. Both utilized automated scoring.

2005 – The LOEP tests were renamed the ESL Tests.

2016 – Five evidence-based ACCUPLACER placement tests were made available to replace those introduced in 1985 and 1990. These are ACCUPLACER Reading; Writing; Arithmetic; Quantitative Reasoning, Algebra, and Statistics; and Advanced Algebra and Functions.

The ACCUPLACER Placement Tests

Rationale for Development

Even though research confirms that ACCUPLACER has always identified students' strengths and weaknesses in the core subjects of reading, writing, and math; over the years, College Board has continued to pay close attention to research and evidence concerning essential prerequisites for college and career readiness and success; changes in academic standards, curriculum, and instruction in the areas of reading, writing, and math; feedback from ACCUPLACER users, including faculty, teachers, counselors, focus groups, advisory panels, and institutions; and feedback from high school and higher education stakeholders who use ACCUPLACER scores and other measures to make course placement decisions.

In 2014, responding to the above input and feedback, we began to design new ACCUPLACER placement tests in a coordinated effort with secondary and postsecondary initiatives to propel all students toward the attainment of both college and career readiness and success. In September 2016, five evidence-based placement

tests were launched: Reading; Writing; Arithmetic; Quantitative Reasoning, Algebra, and Statistics; and Advanced Algebra and Functions. These replace five existing ACCUPLACER placement tests: Reading Comprehension, Writing Skills, Arithmetic, Elementary Algebra, and College-Level Math.

Description of Key Features

In 2014 College Board began to redesign existing ACCUPLACER placement tests to focus on fewer topics that would be most relevant to postsecondary readiness and success. All components in the resulting placement tests align to good classroom instruction, demanding deep thinking and rigorous analysis on questions grounded in real-world knowledge. The features of these ACCUPLACER tests fall into eight categories, listed below. Several of these changes have been integrated into the tests to such an extent that they are key design elements of the revised assessments, as seen in Chapter 3: Test Development Procedures.

Key ACCUPLACER placement test features include:

Text complexity. Text complexity is a measure of passages' inherent reading challenge irrespective of question complexity or difficulty. ACCUPLACER Reading and Writing Tests use a specified range of text complexity aligned to college and career readiness levels of reading. Passages range from "somewhat challenging" to "highly complex," with "complex" reflecting the college and career readiness threshold. (See Appendix B: Text Complexity (Qualitative)—Reading and Writing for details on the qualitative text complexity rubric.)

Words in context. On the ACCUPLACER Reading and Writing Tests, test takers are called on to engage in close reading of texts and to derive the meaning of words and phrases from the contexts in which they are used. The skills and knowledge tested are broadly useful in numerous subject areas and careers. The Reading Test includes questions that assess vocabulary, including word- and phrase-meaning questions, in both extended contexts and single sentences. On both the Reading and Writing Tests, students are also presented with other vocabulary-related challenges, including analyzing word choice rhetorically and improving the precision, concision, and context appropriateness of expression.

Command of evidence. In the ACCUPLACER Suite, test takers analyze material across a wide range of disciplines and content areas (literature and literary nonfiction, science, humanities, and history/social studies) and on career-related topics. They use textual evidence to support their answers and apply an understanding of how authors make use of evidence.

English language conventions. Skilled expression in language requires an understanding of the conventions of standard written English, the developed ability to apply learned language "rules" in the service of not only correctness of expression but also intended rhetorical purposes. The ACCUPLACER Writing Test assesses language conventions and includes a category of questions that assess effective language use in the context of multiparagraph passages that test takers must revise and edit.

Disciplinary literacy. Students' literacy development shouldn't be seen as merely the fostering of generic communication skills, but instead should be grounded in making students familiar with the differing literacy demands of particular fields of study. Some of these differences include specialized vocabulary, types of language used to communicate ideas, text structures and features, and sources of information within and across disciplines. The range of texts included in ACCUPLACER placement tests supports the teaching and assessment of literacy skills across a wide range of disciplines.

Math that matters most. The ACCUPLACER Math Tests are rooted on the philosophy of deeper focus on fewer topics. In keeping with this philosophy, the ACCUPLACER Math Tests focus on skills and knowledge that are essential for college and career readiness.

Applied reasoning questions are emphasized over reasoning questions disconnected from the mathematics curriculum. There is also a strong emphasis on both fluency with mathematical procedures and conceptual understanding.

Problems grounded in real-world contexts. Test takers engage with questions grounded in the real world and directly related to the work performed in college and career. Both the Reading and Writing Tests include literature and literary nonfiction, but they also feature passages that students are likely to encounter in science, social science, and other majors and careers. The Math Tests, which focus on applied reasoning skills essential for college and career readiness, feature multistep applications for solving problems in science, social science, career scenarios, and other real-life contexts.

Connection to math pathways. Because the mathematics prerequisites of college majors and career paths can be vastly different, the three Math Tests deliver a strong connection to the wide variety of course sequences in both STEM and non-STEM fields of study.

The score reporting for ACCUPLACER placement tests is on a scale ranging from 200 to 300. For every test, Skills Insight™ provides a set of data-driven statements for specified score ranges intended to help students, parents, admission officers, educators, and counselors interpret test performance.

Benefits

As mentioned earlier in this document, the ACCUPLACER placement tests are deeply informed by evidence attesting to the essential requirements for college and career readiness and success. The evidentiary base supporting these assessments encompasses College Board's own research, including the results of its national curriculum surveys; state standards with a clear, evidence-based approach to fostering college and career readiness and success for all students; highly regarded scholarly research in relevant fields; feedback from focus groups of educators, including ACCUPLACER users; and other sources.

Because the design of the ACCUPLACER placement tests is informed by the above, they reflect the major shifts taking place in high school instruction, standards, and assessment, and are subsequently able to offer students, parents, instructors, and counselors a better indicator of students' readiness for college and career. Combined with high school grades and other factors, these tests give students an opportunity to demonstrate how well they have attained the skills and knowledge necessary for postsecondary-level work. Other benefits include the following:

Learning, not memorizing. The ACCUPLACER placement tests require students to have a stronger command of fewer topics. Students are asked to apply deep understanding of the skills and concepts most important for college and career readiness and success.

Connection to classroom learning and experience. Students encounter assessments that are closely connected to their classroom experience and that reward focused work and the development of valuable, durable skills and knowledge. The questions and approaches students encounter on the tests are familiar to them because they are modeled on skills and knowledge that matter most for success in college and workforce training according to research and feedback from educational practitioners, some of whom are ACCUPLACER users.

Free resources for practice and review. Students have access to free resources that introduce them to ACCUPLACER and give them a chance to enhance their preparation with targeted review and practice. For more information about free ACCUPLACER resources, visit the ACCUPLACER practice website, **https://accuplacer.collegeboard.org/student/practice**.

ACCUPLACER ESL Assessments

Rationale for Development

As previously stated, findings from a nationwide survey revealed that two-year colleges had a need for an ESL placement assessment. Specifically, they needed tests measuring the language skills of ELLs whose proficiency levels might be too low to allow them to achieve meaningful or actionable scores on the classic ACCUPLACER Reading Comprehension and Sentence Skills placement tests. The Levels of English Proficiency (LOEP) Tests, renamed English as a Second Language Tests in 2005, were developed to respond to this need.

The primary purpose of ACCUPLACER ESL tests—Reading Skills, Sentence Meaning, Language Use, and Listening—is to support the placement of ELLs into English language courses so that they may receive appropriate and targeted levels of instruction. Courses into which students could be placed on the basis of ESL performance include ESL and developmental courses in reading, language arts, and English.[2]

Description of Key Features

ACCUPLACER ESL assesses test takers' command of the literacy skills needed to acquire academic content taught in English-speaking institutions and their ability to communicate in English, both in an academic environment and in everyday situations that ELLs are expected to encounter.

Key ACCUPLACER ESL test features include:

Skills and knowledge that matter most. Questions in ACCUPLACER ESL focus on the language skills and knowledge that ELLs need in order to be successful. Test takers are assessed on their reading comprehension skills, such as the ability to identify and locate information that is explicitly stated as well as infer from contextual clues information that isn't directly stated. Some questions measure the knowledge of grammatical conventions students need in order to express intended meaning and comprehend texts, utterances, or conversations in a range of situations and on a range of subject areas.

Accessible contexts. As tests for students representing a wide range of cultural, educational, and linguistic backgrounds; experiences; and lengths of exposure to English, English-speaking environments, and U.S. culture, care is taken to include contexts on the ESL tests that are as familiar to as wide a range of students as possible. Prior topic-specific knowledge, including knowledge of U.S. culture or norms, isn't tested.

Accessible language. While language isn't simplified when it's part of the construct being assessed, care is taken to make sure passages and questions use language that is as clear and unambiguous as possible. For example, unduly challenging words that aren't part of the construct tested are avoided, and the use of negatives and constructions utilizing "not" in the questions' stems and options is avoided, as they can cause confusion and prevent test takers from demonstrating the knowledge and skills the questions are designed to measure.

Benefits

By developing tests that focus on the core competencies that college-bound ELLs should have or need to acquire in order to succeed in college, the ACCUPLACER ESL Tests offer students, parents, instructors, and counselors a strong indicator of the level of English language instruction these test takers would need to allow them to be successful in postsecondary work in the United States.

Students have access to free resources that introduce them to the ACCUPLACER ESL assessments and give them a chance to enhance their review and preparation. For more information about free ACCUPLACER resources, visit the ACCUPLACER practice website, **https://accuplacer.collegeboard.org/student/practice**.

---

2. In most cases, placement into college-level courses will be based on the ACCUPLACER placement Reading and Writing Tests rather than on ESL.

WritePlacer

Rationale for Development

WritePlacer and WritePlacer ESL, the essay tests in the ACCUPLACER Suite, were developed to support institutions that require actual student writing samples as part of their placement process. The test taker is first administered a prompt that presents the topic and writing task. In response, the test taker demonstrates the extent to which he or she is able to construct a multiparagraph essay that states a position and supports its merit. WritePlacer and WritePlacer ESL scores can be used in conjunction with other ACCUPLACER test scores to assist with the placement of students into college-level, developmental, or various levels of ESL courses.

Description of Key Features

WritePlacer and WritePlacer ESL are designed to allow test takers to consider a given topic, draw on their own ideas and experiences, and present a discussion of their personal positions or perspectives. Because test takers come from diverse backgrounds and as the goal of the essay tests is to elicit viable writing samples that can be evaluated for college placement purposes or for purposes of providing further and appropriate assistance, the two WritePlacer tests have distinct features. These include:

Skills and knowledge that matter most. The holistic rubrics developed for WritePlacer and WritePlacer ESL delineate the skills and knowledge judged by writing experts and practitioners to be critical for the target group of students. The scoring rubric, or evaluation criteria, for WritePlacer evaluates the extent to which the writer addresses the issue presented in the prompt by maintaining purpose and focus; develops, orders, and connects ideas; and demonstrates control of vocabulary, voice, structure, and standard English conventions. Responses on the WritePlacer ESL are evaluated on dimensions deemed by ESL educators to be critical for the success of ELLs: the extent to which writers are able to develop a response to a prompt by using a wide range of words and sentence patterns and by demonstrating command of grammatically correct English.

Accessible contexts. As WritePlacer and WritePlacer ESL are taken by students from a wide range of backgrounds and experiences, it's critical that the prompts used to focus the writing task and elicit writing samples are accessible. For this reason, prompts are (by design) short, simply stated, culturally agnostic, and, to the extent practicable, free of any barriers that could prevent test takers from constructing scorable essays. So that they are accessible to as wide a range of students as possible, prompts for both tests don't require substantive use of textual evidence. Prior topic-specific knowledge, including knowledge of U.S. culture or norms, isn't tested.

Benefits

By developing essay tests that focus on the core writing competencies college-bound students should have or need to further develop in order to succeed in college, WritePlacer and WritePlacer ESL offer students, parents, instructors, and counselors a better indicator to help facilitate test takers' path to readiness for and success in postsecondary work in the United States.

Students have access to free resources that introduce them to the ACCUPLACER assessments and give them a chance to enhance their review and preparation. For students taking the essay tests, College Board provides the WritePlacer and WritePlacer ESL Guides. These guides give students a preview of authentic prompts, sample responses that exemplify each score point, and annotations explaining why a score is received, so that students may gauge where they are based on their performance on the test and practice in order to attain needed proficiency. To access WritePlacer Guides, visit the ACCUPLACER practice website at **https://accuplacer.collegeboard.org/student/practice**.

The Role of Rigorous High School Coursework in ACCUPLACER Success

The academic rigor of a student's high school experience is an important component of success in college and career. Unfortunately, far too many of our high school graduates are leaving K–12 education without the knowledge and skills they need to enter and succeed in some form of postsecondary education, which encompasses the majority of workforce training programs. Students spend far too much time in school focusing their daily work on texts that aren't sufficiently complex instead of focusing on evaluating evidence in reading and writing or on the command of math that matters the most to their future. They aren't practicing the core work that matters the most in every course that they take. They face barriers that few of them overcome in acquiring the skills and knowledge needed to complete postsecondary education. As stated earlier in this manual, one of the foundations of the ACCUPLACER placement tests is a greater connection to classroom learning and experience, so that students can acquire the knowledge that truly matters for college and career.

We must promote greater access to higher education by proactively working with middle schools and high schools to encourage all students to take the right number of core courses and to take the right kinds of courses.

Studies have demonstrated that students who take more challenging coursework in high school are more likely to be ready for college and career by the time they graduate from high school than students who take less challenging coursework. There is a strong relationship between challenging coursework in high school and HSGPA, college enrollment, first-year college grade point average, and persistence to a second year of college (Wyatt, Kobrin, Wiley, Camara, & Proestler, 2011).

We must promote greater access to challenging courses in high school by all students, particularly low-income, underrepresented students. Providing opportunities for access to challenging coursework in high school will increase equity in education and better prepare all students to succeed in college and career (Wyatt, Wiley, Camara, & Proestler, 2011).

In addition to offering and delivering challenging coursework, high schools play an important role in creating a college-going culture for students; students who attend high schools where there are high expectations, strong support for college attendance, and high participation in financial aid applications are more likely to apply, be accepted, and enroll in colleges that match their qualifications (Roderick, Coca, & Nagoka, 2011).

Improving college readiness can address the issue of inequity in education by increasing college graduation rates for all students, regardless of their ethnicity or household income levels. Students face inequitable access to academic rigor in high school, especially underrepresented minority students and low-income students (College Board, 2014). Providing access to rigor increases equity in education and better prepares students for success in college and career (Barry & Niu, 2013).

The Importance of Test Practice and ACCUPLACER Success

AERA/APA/NCME *Standards for Educational and Psychological Testing,* Standard 8.1, states:

> *Information about test content and purposes that is available to any test taker prior to testing should be available to all test takers. Shared information should be available free of charge and in accessible formats (AERA, APA, & NCME, 2014, p. 133).*

In keeping with this standard and College Board's previously stated belief in providing the same access to information and opportunities to all test takers, all ACCUPLACER test takers are given access to free study and review resources. These resources, designed to help students identify and fill in skills and knowledge gaps prior to testing, include sample and practice questions for multiple-choice tests written or reviewed by College Board and are available in accessible formats; sample items are available in ADA-compliant PDF formats, and the ACCUPLACER Study App allows students to not only preview the design and format of a test, but also experience responding to questions on a computer. Students

taking the essay tests have access to free guides that provide test information and two sample essays for each of the available score points along with annotations that explain why the essay was given the indicated score. Free ACCUPLACER review and practice resources are available here: **https://accuplacer.collegeboard.org/student/practice**.

This access to free practice and review materials achieves two key goals. First, these resources give students who use them a chance to demonstrate what they have learned and can do. More importantly, by providing access to these materials and giving all our students, including low-income, underrepresented and underserved students, and nontraditional students, an opportunity to achieve, we uphold College Board's pledge to adhere to the standards for test takers' rights. Standard 8.0 in *The Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 2014, p. 133) states:

> *Test takers have the right to adequate information to help them properly prepare for a test so that the test results accurately reflect their standing on the construct being assessed and lead to fair and accurate score interpretations (AERA, APA, & NCME, 2014, p. 133).*

1.3 Description of Content

The ACCUPLACER Suite of Assessments' test domain definitions are based on the highest-quality information and resources available about the essential requirements for college and career readiness and success. Scholarly research and empirical data derived from curriculum surveys conducted by College Board and other organizations play an important role in informing these definitions. Our staff works with educational experts in examining the evidence and defining the domain of skills and knowledge to be measured in accordance with each assessment's primary purpose and the claims associated with each assessment. We also prepare test and question/task specifications that represent the depth and breadth of the defined domains and help ensure the consistent development of assessments of the highest quality. This section provides an overview of the content of the tests and their question and task formats, in keeping with standards and best practices (AERA, APA, & NCME, 2014). Table 1.2 presents an overview of the test format for the ACCUPLACER Suite. For a more in-depth look at test content and specifications, see Chapter 3: Test Development Procedures. For the psychometric properties of the tests, see Chapter 6: Psychometrics.

Table 1.2: *Number of Questions for the ACCUPLACER Suite of Assessments*

Test	Number of Questions		
	Discrete	Set-Based	Total
Reading	12	8	20
Writing	–	25	25
Arithmetic	20	–	20
Quantitative Reasoning, Algebra, and Statistics	20	–	20
Advanced Algebra and Functions	20	–	20
ESL Reading Skills	6	14	20
ESL Sentence Meaning	20	–	20
ESL Language Use	20	–	20
ESL Listening	20	–	20
WritePlacer	1 essay		1
WritePlacer ESL	1 essay		1

ACCUPLACER Placement Tests

ACCUPLACER Reading Test

The ACCUPLACER Reading Test is a computer-adaptive assessment of test takers' developed ability to derive meaning from a range of prose texts and to determine the meaning of words and phrases in short and extended contexts. The test comprises 20 questions and is intended to collect evidence in support of a broad claim about student performance:

> *Students can demonstrate college and career readiness proficiency in reading and comprehending a broad range of appropriately challenging literary and informational texts in the content areas of literature and literary nonfiction, careers/history/social studies, humanities, and science.*

Passages on the test cover a range of content areas (including literature and literary nonfiction, careers/history/social studies, humanities, and science), writing modes (informative/explanatory, argument, and narrative), and complexities (relatively easy to very challenging). Both single and paired passages are included. The test pool includes both authentic texts (previously published passages excerpted or minimally adapted from their published form) and commissioned texts (written specifically for the test). Passages, which may be previously published or written for the test, represent two basic genres: literary or informational.

Questions are multiple-choice in format and ask test takers to derive information and ideas (stated or implied) from passages. The questions may be discrete (standalone) or part of sets built around a common passage or passages. Four broad knowledge and skill categories are assessed: Information and Ideas (reading closely, determining central ideas and themes, summarizing, understanding relationships); Rhetoric (analyzing passages rhetorically in terms of such features as word choice, text structure, point of view, purpose, and argumentation); Synthesis (analyzing and drawing connections between pairs of related texts); and Vocabulary (determining word and phrase meanings in context).

For a more in-depth look at test content and specifications, see Chapter 3: Test Development Procedures.

Quick Facts:

- The computer-adaptive Reading Test has 20 questions; the linear, accommodated COMPANION test has 40 questions.
- All questions are multiple-choice.
- Some questions are based on passages; others are sentence-based.
- Questions may be discrete or set-based.
- One total Reading Test score, ranging from 200 to 300, is reported.

ACCUPLACER Writing Placement Test

The ACCUPLACER Writing Test is a computer-adaptive assessment of test takers' developed ability to revise and edit a range of prose texts for effective expression of ideas and for conformity to the conventions of Standard Written English sentence structure, usage, and punctuation. The test comprises 25-questions and is intended to collect evidence in support of a broad claim about student performance:

> *Students can demonstrate college and career readiness proficiency in revising and editing a range of texts in a variety of content areas, both academic and career related, for expression of ideas and for conformity to the conventions of standard written English grammar, usage, and punctuation.*

A series of extended prose passages and their associated multiple-choice questions constitute the Writing Test. Passages cover a range of content areas (including literary nonfiction, careers/history/social studies, humanities, and science), writing modes (informative/explanatory, argument, and narrative), and complexities (relatively easy to

very challenging). All passages are commissioned—that is, written specifically for the test—so that "errors" (a collective term for a wide range of rhetorical and conventions-related problems) can more effectively be introduced into them.

Questions are multiple-choice in format and appear in sets built around a common, extended passage; no discrete (standalone) questions are included in the Writing Test. In answering the questions, test takers must determine the best revision or editing decision in a particular case (or that no change should be made to the passage as originally presented).

For a more in-depth look at test content and specifications, see Chapter 3: Test Development Procedures.

Quick Facts:

- The computer-adaptive Writing Test has 25 questions; the linear, accommodated COMPANION test has 40 questions.
- All questions are multiple-choice.
- Questions are based on passages.
- One total Writing Test score, ranging from 200 to 300, is reported.

ACCUPLACER Arithmetic Placement Test

The ACCUPLACER Arithmetic Test is a computer-adaptive test. The test comprises 20 questions and is intended to collect evidence in support of the following claim about student performance:

Students can demonstrate fluency with, understanding of, and the ability to apply fundamental arithmetic concepts, skills, and practices that show progress toward success in college courses, career training, and career opportunities.

The Arithmetic Test assesses five knowledge and skill categories: Whole number operations; Fraction operations; Decimal operations; Percent; and Number comparisons and equivalents. Questions on the Arithmetic Test focus on computation, order of operations, estimation and rounding, comparing and ordering values in different formats, and recognizing equivalent values across formats. Test takers need to exhibit command of mathematical practices, fluency with mathematical procedures, and conceptual understanding of mathematical ideas. The problems explore the full dynamic range of each content area through precisely crafted questions that emphasize the use of math in unlocking insights and solving problems. All questions are multiple-choice.

For a more in-depth look at test content and specifications, see Chapter 3: Test Development Procedures.

Quick Facts:

- The computer-adaptive Arithmetic Test has 20 questions; each of the two linear, accommodated COMPANION tests has 40 questions.
- All questions are multiple-choice, administered in a discrete question format.
- Some questions present test takers with problem-solving scenarios.
- Some questions allow for the use of a four-function calculator.
- One total Arithmetic Test score, ranging from 200 to 300, is reported.

ACCUPLACER Quantitative Reasoning, Algebra, and Statistics (QAS) Placement Test

The ACCUPLACER QAS Test is a computer-adaptive test. The test comprises 20 questions and is intended to collect evidence in support of the following claim about student performance:

> *Students can demonstrate fluency with, understanding of, and the ability to apply the mathematical concepts, skills, and practices that are most strongly prerequisite and central to their ability to progress through a range of college courses, career training, and career opportunities.*

The QAS Test assesses test takers' developed ability for mathematics content suited for students entering many non-STEM fields of study or for students who are undecided on a major. Questions focus on a range of knowledge and skill categories, including Rational numbers; Ratio and proportional relationships; Exponents; Algebraic expressions; Linear equations; Linear applications and graphs; Probability and sets; Descriptive statistics; Geometry concepts for Pre-Algebra; Geometry concepts for Algebra 1. All questions are multiple-choice.

For a more in-depth look at test content and specifications, see Chapter 3: Test Development Procedures.

Quick Facts:

- The computer-adaptive QAS Test has 20 questions; the linear, accommodated COMPANION test has 40 questions.
- All questions are multiple-choice, administered in a discrete question format.
- Some questions present test takers with problem-solving scenarios.
- Basic four-function and square root calculators are allowed on some questions.
- One total QAS Test score, ranging from 200 to 300, is reported.

ACCUPLACER Advanced Algebra and Functions (AAF) Placement Test

The ACCUPLACER AAF Test is a computer-adaptive test. The test comprises 20 questions and is intended to collect evidence in support of the following claim about student performance:

> *Students can demonstrate fluency with, understanding of, and the ability to apply more advanced mathematical concepts, skills, and practices that are most strongly prerequisite and central to their ability to progress through a range of college courses, career training, and career opportunities.*

The AAF Test assesses test takers' developed ability for mathematics content suited for students entering STEM fields of study, as well as students entering non-STEM fields of study that require some advanced math (e.g., medicine, economics, accounting). The knowledge and skill categories assessed include Linear equations; Linear applications and graphs; Factoring; Quadratics; Functions; Radical and rational equations; Polynomial equations; Exponential and logarithmic equations; Geometry concepts for Algebra 1; Geometry concepts for Algebra 2; and Trigonometry. All questions are multiple-choice.

For a more in-depth look at test content and specifications, see Chapter 3: Test Development Procedures.

Quick Facts:

- The computer-adaptive AAF Test has 20 questions; the linear, accommodated COMPANION test has 40 questions.
- All questions are multiple-choice, administered in a discrete question format.
- Some questions present test takers with problem-solving scenarios.
- Basic, square root, and graphing calculators are allowed on some questions.
- One total AAF Test score, ranging from 200 to 300, is reported.

English as a Second Language Tests

The ACCUPLACER ESL assessments comprise four multiple-choice tests and one essay test: ESL Reading Skills, ESL Sentence Meaning, ESL Language Use, ESL Listening, and WritePlacer ESL, respectively.

ESL Reading Skills Test

The ESL Reading Skills test is a 20-question, computer-adaptive test that is intended to collect evidence in support of the following claim about student performance:

> ELLs can demonstrate their developed ability to read and comprehend appropriately challenging short passages and narratives in a variety of content areas by referring to what the texts say explicitly and drawing reasonable inferences from the passages to compensate for gaps in their lexical and structural control of the language.

Half of the questions on the ESL Reading Skills Test assess explicitly stated information, and the other half measure test takers' developed ability to infer information not directly stated in a passage. All questions are multiple-choice.

For a more in-depth look at test content and specifications, see Chapter 3: Test Development Procedures.

Quick Facts:

- The computer-adaptive ESL Reading Skills Test has 20 questions; each of the two linear, accommodated COMPANION tests has 35 questions.
- All questions are multiple-choice.
- Questions are based on passages.
- Questions may be discrete or set-based.
- One placement test score, ranging from 20 to 120, is reported.

ESL Sentence Meaning Test

The ESL Sentence Meaning Test is a 20-question, computer-adaptive test that is intended to collect evidence in support of the following claim about student performance:

> ELLs can demonstrate their developed ability to read and comprehend sentences written in a variety of structures on a variety of topics, compensating for gaps in their lexical and structural control of the language by using contextual clues.

Sentences selected for use as stimuli on the test are drawn from the content areas of natural science, history/social studies, arts/humanities, psychology/human relations, and practical situations. Questions test takers are asked to answer cover four content areas: particles, phrasal verbs, and prepositions; adverbs, adjectives, and connectives sequences; basic nouns and verbs; and commonly used Idioms. All questions are multiple-choice.

For a more in-depth look at test content and specifications, see Chapter 3: Test Development Procedures.

Quick Facts:

- The computer-adaptive ESL Sentence Meaning Test has 20 questions; each of the two linear, accommodated COMPANION tests has 35 questions.
- All questions are multiple-choice.
- Questions are administered in a discrete format.
- One placement test score, ranging from 20 to 120, is reported.

ESL Language Use

The ESL Language Use Test is a 20-question, computer-adaptive test that is intended to collect evidence in support of the following claim about student performance:

> *ELLs can demonstrate their developed ability to identify and apply learned conventions of standard written English grammar, usage, and punctuation to a range of English sentence structures on a variety of topics.*

Specific grammatical usage conventions measured are nouns, pronouns, and pronoun case structure; sentence structure; subject-verb agreement; adverbs/adjectives; verbs; and subordination/coordination. All questions are multiple-choice. Some require students to choose a word or phrase to fill in a blank, while others ask students to choose the best way to combine two sentences.

For a more in-depth look at test content and specifications, see Chapter 3: Test Development Procedures.

Quick Facts:

- The computer-adaptive ESL Language Use Test has 20 questions; each of the two linear, accommodated COMPANION tests has 35 questions.
- All questions are multiple-choice.
- All questions are administered in a discrete question format.
- One placement test score, ranging from 20 to 120, is reported.

ESL Listening Test

The ESL Listening Test is a 20-question, computer-adaptive test that is intended to collect evidence in support of the following claim about student performance:

> *ELLs can demonstrate their developed ability to understand ideas and information presented in connected discourse on a variety of topics, compensating for gaps in their lexical and structural control of the language by using contextual clues.*

The ESL Listening Test is a direct measure of the listening skills of test takers who are nonnative speakers of English. Students don't have to be able to read English to successfully complete this test. Instead, it assesses test takers' developed ability to listen to and understand one or more people speaking in English through a series of conversations or lectures. The conversations take place in academic environments as well as everyday situations that ELLs might expect to experience. Questions assess literal comprehension of information delivered via such connected speech as well as information and ideas that are implied.

For a more in-depth look at test content and specifications, see Chapter 3: Test Development Procedures.

Quick Facts:

- The computer-adaptive ESL Listening Test has 20 questions; COMPANION versions with content that would be read aren't made available, as these would violate the construct being assessed.
- All questions are multiple-choice.
- Questions, based on conversations or lectures, are set-based.
- One placement test score, ranging from 20 to 120, is reported.

ACCUPLACER Essay Tests

In addition to the multiple-choice tests, ACCUPLACER offers WritePlacer and WritePlacer ESL, both essay tests that provide direct measures of the writing skills of test takers. Both require students to write an original essay ranging from 300 to 600 words.

WritePlacer

As with the other tests in the ACCUPLACER Suite, WritePlacer is intended to collect evidence in support of a broad claim about student performance:

Students can demonstrate their developed ability to successfully create an original essay in which they use reasoning, personal experience, observations, readings, and an appropriate rhetorical approach to effectively develop a point of view or position on an issue.

WritePlacer assesses the writing skills of entering college students. Test scores help colleges determine whether a student is ready for college-level coursework.

Students taking WritePlacer are presented with a prompt and asked to write an essay of 300 to 600 words. A prompt consists of a short passage and an assignment that requires the student to focus on the issue addressed in the passage. WritePlacer measures a student's developed ability to do the kind of writing required in most college courses: writing that emphasizes precise use of language, logical presentation of ideas, development of a point of view, and clarity of expression.

Test taker responses, scored on a holistic rubric, are judged on six dimensions: purpose and focus; organization and structure; development and support; sentence variety and style; mechanical conventions; and critical thinking.

WritePlacer essays are electronically scored, and scores range from 1 to 8. An essay that is too short to be evaluated, written on a topic other than the one presented, or written in a language other than English will be given a score of zero.

For a more in-depth look at test content and specifications, see Chapter 3: Test Development Procedures.

Quick Facts:

- WritePlacer is administered on the computer; two accommodated COMPANION forms are available.
- Essays are scored electronically.
- One placement test score, ranging from 1 to 8, is reported.

WritePlacer ESL

WritePlacer ESL assesses the writing skills of college-bound ELLs. It's intended to collect evidence in support of a broad claim about student performance:

ELLs can demonstrate their developed ability to successfully create an original essay in which they draw on a broad range of experiences, learning, and ideas to effectively develop a point a view or position on a given topic.

Test takers are presented with a prompt and asked to write an essay. WritePlacer ESL prompts are carefully designed to be culturally accessible to English language learners. They are short, free of technical or specific literary references, and don't require specialized topical or cultural knowledge. Students are asked to respond to the prompt by drawing on a broad range of experiences, learning, and ideas.

Test taker responses, scored on a holistic rubric, are judged on observed ability to organize and develop an essay using appropriate words, sentences, and grammatical conventions.

WritePlacer ESL essays are electronically scored, and scores range from 1 to 6. An essay that is too short to be evaluated, written on a topic other than the one presented, or written in a language other than English will be given a score of zero. Test scores help colleges determine the level of English instruction ELL students should receive so that they may succeed in credit-bearing courses.

For a more in-depth look at test content and specifications, see Chapter 3: Test Development Procedures.

Quick Facts:

- WritePlacer ESL is administered on the computer; two accommodated COMPANION forms are available.
- Essays are scored electronically.
- One placement test score, ranging from 1 to 6, is reported.

CHAPTER 2

Fairness

Fairness is a central tenet of all College Board assessments, one with a number of wide-ranging implications. This chapter begins (Section 2.1) with an overview of the aspects of fairness that are considered in the development and administration of a College Board assessment. The chapter then looks at the constructs, purposes, and population relevant to the ACCUPLACER Suite of Assessments (Section 2.2), because test fairness cannot be evaluated separately from an understanding of what a given test is purporting to measure, what purpose or purposes it's intended to serve, and who comprises the test-taking population. The chapter then provides an in-depth look (Section 2.3) at the steps that are taken to ensure fairness in all College Board assessments, including the ACCUPLACER Suite, from the initial test design and development, through test administration and scoring interpretation. The chapter then concludes (Section 2.4) by putting into focus the importance of fairness and College Board's commitment to provide it to all test takers.

2.1 Fairness in College Board assessments

College Board believes in providing all test takers with a fair opportunity to demonstrate what they know and are able to do. Conceptually, fairness can be defined in terms of both equitable treatment of all test takers in test administration and equal measurement quality across subgroups and populations. Best practices and standards 3.1–3.5 of the AERA/APA/NCME *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, APA, & NCME, 2014, p. 63). An assessment should be built in such a way that the constructs being assessed are measured equally for all intended test takers and test-taking subgroups, and it should be administered in a manner that is fair and equitable for all test takers, regardless of gender, race/ethnicity, and other characteristics.

In order to accomplish these goals, four aspects of fairness, identified by the AERA/APA/NCME Standards, should be addressed when developing and administering an assessment.

1. **Fairness in treatment during the testing process.** Fairness in treatment involves "[maximizing], to the extent possible, the opportunity for test takers to demonstrate their standing on the construct(s) the test is intended to measure" (AREA/APA/NCME, 2014, 51). The AERA/APA/NCME 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 also that they increasingly recognize that "sometimes flexibility is needed to provide essentially equivalent opportunities for some test takers" (AERA, APA, & NCME, 2014, 51) when accommodations (modifications) in testing don't compromise the construct being measured (e.g., reading achievement).

2. **Fairness as lack of measurement bias.** Per the AERA/APA/NCME Standards, bias in a measurement itself or in the predictions made from it may occur when "characteristics of the test itself that aren't 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" (AERA, APA, & NCME, 2014, 51). Bias in this sense can play out as differential performance on items and/or tests by identified subgroups equally matched on the characteristic of interest (e.g., mathematics achievement) and/or in differential predictions (inferences) about such subgroups. It's a responsibility of test makers to identify and root out such construct-irrelevant factors when they advantage or disadvantage subgroups of test takers.

3. **Fairness in access to the construct(s) being measured.** The AERA/APA/NCME 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" (AERA, APA, & NCME, 2014, p. 52). Accommodations may take such forms as providing students with visual impairments access to large-print versions of text (when visual acuity isn't the construct being measured) and avoiding the use of regional expressions in test questions intended for a national or international audience.

4. **Fairness as validity of individual test score interpretations for the intended uses.** The AERA/APA/NCME Standards indicate that test makers and users should attend to differences among individuals when interpreting test data and not generalize about individuals from the performance of subgroups to which they belong. Given those considerations, "adaptations to individual characteristics [e.g., reducing language barriers in testing when language proficiency isn't the construct being measured] 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" (AERA, APA, & NCME 2014, p. 53–54).

College Board embraces the fairness guidelines articulated by the AERA/APA/NCME Standards and the overarching goal of ensuring the maximal inclusiveness, representativeness, and accessibility of its test materials consistent with the designs and aims of the tests. Through its fairness-related documentation, processes, procedures, staff and consultant trainings, and support materials, College Board strives to ensure that all its tests, including tests in the ACCUPLACER Suite:

- are appropriate for and accessible to a national and international test-taking population, and defined subgroups of that population, taking a medium- to high-stakes assessment of college and career readiness;

- neither advantage nor disadvantage individual test takers or defined population subgroups of test takers due to factors not related to the constructs (e.g., reading comprehension, mathematics achievement) being measured;

- are free of content or contexts likely to give offense, provoke a strongly distracting emotional response, or otherwise inhibit test takers from performing their best work on the tests;

- accurately and fairly portray the diverse peoples of the United States and the world and convey the widest possible range of ideas, perspectives, and experiences consistent with the tests' designs;

- make test content as fully and as widely accessible to as many test takers as possible through design and development processes yielding materials consistent with the principles of universal design and through a range of accommodations for test takers with particular needs, while remaining faithful, to the fullest extent possible, to the constructs being measured;

- have clearly articulated purposes and uses for which they and their data should and shouldn't be used and have clearly indicated populations for whom the tests are designed.

This chapter provides a brief overview of several interrelated issues of fairness as they pertain to the ACCUPLACER Suite assessments—specifically, how College Board ensures the fairness of its test content, makes that content as accessible as possible,

and provides accommodations for students needing them. Preceding this discussion is a concise description of the ACCUPLACER Suite intended to provide needed context for evaluating College Board's test fairness practices.

2.2 ACCUPLACER Suite constructs, purposes, and populations

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

- The "ACCUPLACER Suite of Assessments" is College Board's collective term for the college placement, English as a Second Language (ESL), WritePlacer, and WritePlacer ESL tests (presented in Table 1.1 in Chapter 1). The ACCUPLACER placement tests (Reading; Writing; Arithmetic; Quantitative Reasoning, Algebra, and Statistics (QAS); and Advanced Algebra and Functions (AAF)) are typically administered to entering college students to determine the degree to which they are prepared to succeed in college and workforce training programs. They are also administered to high school students who are eligible for dual enrollment either late in the junior year or early in the senior year to evaluate their college and career readiness. Students whose test performance doesn't meet their institution's readiness requirements are placed in developmental or credit-bearing corequisite programs so that they may develop the skills and knowledge needed to succeed in college.

- The ACCUPLACER ESL tests, including WritePlacer ESL, are administered to college-bound English Language Learners (ELLs). They support ELLs' placement into appropriate English language courses, where they may receive the instruction that prepares them for participating successfully in college-level courses.

- WritePlacer is administered to students at institutions that require actual student writing samples as part of their placement process. Test scores may be used in conjunction with other ACCUPLACER test scores to assist with the placement of students into college-level or developmental or credit-bearing corequisite writing courses.

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

- Reading: Students can demonstrate college and career readiness proficiency in reading and comprehending a broad range of appropriately challenging literary and informational texts in the content areas of literature and literary nonfiction, careers/history/social studies, humanities, and science.

- Writing: Students can demonstrate college and career readiness proficiency in revising and editing a range of texts in a variety of content areas, both academic and career related, for expression of ideas and for conformity to the conventions of standard written English grammar, usage, and punctuation.

- Arithmetic: Students can demonstrate fluency with, understanding of, and the ability to apply fundamental arithmetic concepts, skills, and practices that show progress toward success in college courses, career training, and career opportunities.

- Quantitative Reasoning, Algebra, and Statistics: Students can demonstrate fluency with, understanding of, and the ability to apply the mathematical concepts, skills, and practices that are most strongly prerequisite and central to their ability to progress through a range of college courses, career training, and career opportunities.

- Advanced Algebra and Functions: Students can demonstrate fluency with, understanding of, and the ability to apply more advanced mathematical concepts, skills, and practices that are most strongly prerequisite and central to their ability to progress through a range of college courses, career training, and career opportunities.

- ESL Reading Skills: ELLs can demonstrate their developed ability to read and comprehend appropriately challenging short passages and narratives in a variety of content areas by referring to what the texts say explicitly and drawing reasonable inferences from the passages to compensate for gaps in their lexical and structural control of the language.

- ESL Sentence Meaning: ELLs can demonstrate their developed ability to read and comprehend sentences written in a variety of structures on a variety of topics, compensating for gaps in their lexical and structural control of the language by using contextual clues.

- ESL Language Use: ELLs can demonstrate their developed ability to identify and apply learned conventions of standard written English grammar, usage, and punctuation to a range of English sentence structures on a variety of topics.

- ESL Listening: ELLs can demonstrate their developed ability to understand ideas and information presented in connected discourse on a variety of topics, compensating for gaps in their lexical and structural control of the language by using contextual clues.

- WritePlacer: Students can demonstrate their developed ability to successfully create an original essay in which they use reasoning, personal experience, observations, readings, and an appropriate rhetorical approach to develop and support a point of view or position on an issue.

- WritePlacer ESL: ELLs can demonstrate their developed ability to successfully create an original essay in which they draw on a broad range of experiences, learning, and ideas to effectively develop and support a point a view or position on a given topic.

The tests in the ACCUPLACER Suite have two primary purposes: to determine the degree to which students are prepared to succeed in college and in workforce training programs (ACCUPLACER placement tests and WritePlacer) and to ascertain the English language support required for ELLs to achieve the above goal (ACCUPLACER ESL tests, including WritePlacer ESL). In the design of the tests and development of test content, care is taken to ensure the tests in the ACCUPLACER Suite align with these goals. To ensure resulting test scores provide meaningful information about a student in each of the areas tested, each test is designed to collect evidence from student performance in support of a broad claim about what students know and can do, and each claim is aligned to the primary purpose of each of the tests in the suite. ACCUPLACER Suite results should be used with other factors, such as HSGPA, when making high-stakes decisions.

The ACCUPLACER Suite provides data that are used for many purposes by different users. The three key users are higher education, K–12 educators, and students. In keeping with best practices and the requirements outlined in the AERA/APA/NCME

Standards, the ACCUPLACER Suite's primary intended uses and interpretations for each group of primary users are discussed in the following paragraphs, with a rationale presented for each of the three uses:

- Evaluating and monitoring students' college and career readiness
- Making college placement decisions
- Contributing to English language instructional supports for ELLs.

Evaluating and monitoring students' college and career readiness. ACCUPLACER is used by both high school and postsecondary educators and students. For incoming college students, ACCUPLACER placement test scores, along with their corresponding Skills Insight—and, in the case of WritePlacer, along with detailed dimensional score description—serve as meaningful indicators of their readiness for college and career training. For high school students, the ACCUPLACER performance data mentioned above can provide information about what college-level classes these students are ready for in high school and courses for which they need to seek additional supports; states, districts, and schools may use these data to monitor what proportion of their student body has a high likelihood of success in college-entry coursework. For both groups of students, it should be emphasized that while ACCUPLACER test scores and score descriptors, when used with factors such as a student's HSGPA, can enable institutions to make more informed course placement decisions, they aren't intended for decisions that may impact a student's advancement opportunities (e.g., decisions that restrict student access to challenging coursework or discourage aspirations of attaining higher education).

Making college course placement decisions. The ACCUPLACER placement tests, including WritePlacer, are intended for use in higher education to provide a better understanding of students' level of preparedness for college-level work so that colleges may make more informed course placement policies and decisions. Such policies and decisions should be verified empirically, with appropriate adjustments made as necessary, in order to promote positive student and institutional outcomes. To this end, College Board provides the Admitted Class Evaluation Service (ACES), a free online validity study service that helps institutions to gather the predictive validity evidence needed to make or improve placement policies and decisions (see **https://aces.collegeboard.org/** for more information). Colleges and universities using scores from College Board assessments (ACCUPLACER, SAT®, AP®, and CLEP®) in their admission or placement policies are encouraged to make use of ACES to verify the appropriateness of their policies, giving them the ability to make informed decisions about confirming or refining these as appropriate to meet institutional needs.

Contributing to English language instructional supports for ELLs. Postsecondary institutions, such as colleges and organizations with English language programs that support ELLs, use test scores from ACCUPLACER ESL tests and their corresponding proficiency statements to determine the level of English language instruction and support these students will receive. Since students who take the ESL tests are typically those whose core English skills wouldn't allow them to perform well on the college placement tests, the tests that place these students into the appropriate ESL courses play a critical role; by allowing these students to develop the skills and knowledge needed for success on the tests that determine placement into college-level courses, they help identify and make it possible for these institutions to deliver the supports needed for scaffolding ELLs' paths to and through college.

2.3 Fairness of the ACCUPLACER Suite
Test construction
College Board has taken and continues to take numerous, exacting steps to establish and maintain the fairness of all its assessments, including those in the ACCUPLACER Suite. These efforts begin with the test design and continue through its ongoing development.

Test design

Key concepts. Fairness in test content starts with a thoughtfully crafted and sharply focused assessment design. Tests in the ACCUPLACER Suite of Assessments were designed and developed over time. Three ESL tests (Reading Skills, Sentence Meaning, and Language Use) were introduced in 1993. Listening and WritePlacer ESL followed in 2002, as did WritePlacer. Beginning in 2014, College Board undertook a comprehensive redesign of the five ACCUPLACER placement tests that had been used since 1985. The resultant placement tests—Reading, Writing, Arithmetic, QAS, and AAF—along with the ESL and WritePlacer tests, make up the current ACCUPLACER Suite. Tests in the ACCUPLACER Suite were purpose-built using the best available evidence in order to, in the case of the placement tests, measure attainment of essential college and career readiness and success prerequisites and, in the case of ACCUPLACER ESL tests, measure the language achievement of English Language Learners (ELLs) and help identify the language supports needed to prepare them for successfully navigating postsecondary coursework where English is the language for acquiring as well as demonstrating learned information. A central tenet of the design philosophy was that the tests would address in depth the core knowledge and skills that evidence shows matter most. Content assessed includes the knowledge of words and phrases as they are used in context; command of evidence; writing skills required for presenting ideas logically, clearly, and correctly; the math concepts, skills, and practices strongly associated with the requirements of a wide range of college majors and workforce training programs ("math that matters most"); problems grounded in real-world contexts; and close reading of texts. In addition, the design of assessment tasks was intended to align with best classroom teaching practices, thereby reducing the distance between assessment and instruction and making the test content more accessible to students.

Evidence gathering and consultation. In the design and development of assessments, College Board undertakes intensive research, convenes advisory committees, and solicits feedback from faculty members and subject matter experts representing a broad cross section of high school and higher education institutions. This prerelease process continues to be augmented by ongoing research as well as feedback from higher education and K–12. Among the most prominent elements are the following:

- **Curriculum survey.** Every three to four years, College Board undertakes and publishes the results from a survey of a nationally representative sample of middle school, high school, and postsecondary instructors. The primary purpose of the survey is to ascertain what skills and knowledge are deemed prerequisite for readiness for and success in common first-year, credit-bearing postsecondary courses across a range of subjects as a check on whether College Board college readiness assessments measure what postsecondary faculty deem critical for incoming students to know and be able to do. From these (and other) data, we make periodic refinements as warranted to our assessments in order to improve alignment to postsecondary faculty expectations and to better represent best instructional practices.

- **Test review committees.** As described in detail below, College Board convenes groups of secondary and postsecondary educators from around the country as independent consultants to review its test materials for both content soundness and fairness. Feedback from these groups is used to refine or remove material deemed problematic for use in its present form from the development process prior to its use with students in an operational setting. In addition to providing actionable information about specific test materials, these committees offer College Board developers ongoing, vital connections to and information about teaching and learning as they are undertaken in classrooms throughout the United States.

- **Academic advisory committees.** When designing new assessments, College Board convenes academic advisory committees in the relevant subject areas. Composed of leading educators, these committees advise the organization on matters relating to educational philosophy, guiding principles, and standards for creating coherence

among the instructional materials, assessments, and professional development programs and services designed to prepare students for college and workforce training success and on policy decisions regarding equity, parity, and access for all students. Among the core duties of each academic advisory committee are to identify and discuss the core skills and knowledge that research has shown to be essential for college and career readiness and success in their discipline; identify different ways in which these skills and knowledge can be cultivated through courses, assessments, and professional development; identify the ways in which integrated approaches to teaching, learning, and measuring these skills and knowledge can be reflected within and across College Board programs; and identify issues of significance in their discipline for guiding overarching instruction and assessment strategies within College Board.

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

Among their virtues, detailed specifications for test materials—stimuli, questions, and the like—help further the goal of test fairness by ensuring that test content, regardless of when or by whom it was developed, meets the requirements of measuring the desired construct(s), aids in achieving the specified purpose(s) of the test, is suitable for the identified testing population and its subgroups, and is highly consistent in substance. In combination with other steps, such as item calibrations and scaling, the use of carefully articulated specifications helps ensure that test takers, regardless of the date on which they take the test or the particular set of test materials they receive, have highly comparable testing experiences and that their performance isn't influenced by unnecessary variability in test materials (e.g., one student receiving a series of easy reading passages and another receiving a series of hard passages). College Board has developed and maintains extensive documentation for use internally and by its development partners in the construction of stimulus materials (e.g., passages, informational graphics) and test questions and tasks. These development guides include discussion of general issues, such as construct definition and identification of testing purposes and population, as well as detailed process guidelines and examples of effective practice.

As feedback in such forms as student performance data, input from independent external reviewers, and comments from higher education are received, refinements to content and statistical specifications and documentation for the purposes of improving the validity, reliability, and fairness of the assessments are occasionally made. When this happens, College Board communicates those changes to test takers and other stakeholders.

Test review guides. College Board also maintains detailed content and fairness review guides for the independent experts it hires to evaluate its test materials. The content of these guides aligns closely with that of the (internal) stimulus material and question/task development guides. The guides include information about the review process, expectations for reviewers, general content and fairness guidelines, and bullet lists of specific considerations. These detailed guides promote fairness not only through explicit guidance about reviewing test content for fairness but also by virtue of standardizing the review process (i.e., ensuring that different reviewers or groups of reviewers approach the review task in similar ways). These guides are periodically reviewed and are updated to reflect any refinements to the test design.

Test development

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

External fairness review. Prior to pretesting, all ACCUPLACER Suite test materials are reviewed by external, independent reviewers who are asked to evaluate these materials according to a set of criteria for (depending on qualifications) content soundness or fairness. As a group, these reviewers are typically active classroom teachers drawn from across the nation, teach at the secondary and postsecondary levels, and are deeply familiar with both the student population (and subgroups) of interest and the nature and purposes of the assessment. At present, fairness reviewers are primarily individuals from African American, Asian American, Latino/a, American Indian/Native American, and mixed racial/ethnic backgrounds. They live and work in different regions of the United States; teach a range of disciplines, including ESL, at different levels (secondary, postsecondary) in different types (e.g., rural, suburban, urban) of institutions. An equal, or approximately equal, number of males and females are represented; individuals identifying as having genders other than male or female, or no gender, are also welcome to participate.

Fairness reviewers are charged with helping to ensure that test stimuli and questions are broadly accessible to the wide-ranging student population that takes the ACCUPLACER Suite, that the materials don't advantage or disadvantage identified subgroups of test takers on factors unrelated to the construct being assessed, and that the texts and topics addressed are appropriate for the audience of secondary students and the occasion of medium- to high-stakes testing. In addition to employing their own professional judgment and expertise, fairness reviewers are directed to apply criteria developed by College Board for the ACCUPLACER Suite assessments. These criteria include both broad-based considerations and those specific to elements of the individual tests (e.g., literature and literary nonfiction passages in Reading). The following summarizes the "fairness framework" used to evaluate materials across the testing programs composing the ACCUPLACER Suite.

Topics to avoid. Topics that are highly sensitive or controversial among various population groups, generally not required by the constructs (e.g., reading comprehension, writing achievement) being measured by the tests, should typically be avoided.

Portrayal - All population groups should be portrayed fairly, authentically, and with respect. No group should be depicted in a demeaning (including self-demeaning) way by test materials.

Stereotyping - Instances of stereotyping should be avoided, whether the stereotype is negative or "positive." Examples of stereotyping include portraying all women as homemakers and cooks; girls as ballerinas and boys as athletes; African Americans as people who live in the inner city or who excel exclusively in sports; Asian Americans as particularly gifted in math; Native Americans as particularly attuned to nature; or older people as feeble, slow, incompetent, or dependent. All groups should be represented as having the broadest possible range of points of view, talents, interests, and aspirations.

Group Identification - For the appropriate terminology for given population groups, it's generally best practice to use the terminology that respective group members themselves prefer. Outmoded group identifications are sometimes acceptable in certain historical or cultural contexts (e.g., *The Journal of Negro History,* National Association for the Advancement of Colored People). Men and women should be identified in comparable ways (e.g., both by first name or both by last name).

Language - Given that different groups of test takers have different rates of exposure to them, foreign words and phrases, slang and dialect, and idiomatic expressions should be avoided in testing materials unless their use is pertinent to the construct.

Ethnocentrism - Test materials shouldn't treat aspects of U.S. or Western political systems, society, cultures, or values as universal when they are specific to those regions.

Regionalisms - Regional differences are often subtle. Referring to "yard sales" or "garage sales," for example, could be confusing for urban test takers. (Similarly, "stoop sales" could be confusing to rural or suburban test takers.) Generic terms are, therefore, generally preferable to region-specific ones. For example, a term such as "soft drink" is usually preferred over "pop," "soda," or "Coke." On the other hand, reviewers shouldn't make overly restrictive assumptions about test takers' experiences or ability to make inferences from text. Even test takers who have never seen snow, lightning bugs, or a subway, for example, are likely to understand these concepts if the test materials explain them well enough.

Testing Context - The testing situation can be stressful for test takers. Therefore, special care should be taken to avoid content and contexts that seem likely to trigger or add to that stress. Highly controversial topics (such as abortion or the death penalty) that might potentially be discussed safely in a classroom environment, in which a teacher can mediate and students have the time and space to express their ideas and feelings, are often inappropriate for a testing context, during which students are timed and have no emotional outlet.

The foregoing general principles are clarified and refined for reviewers (and internal staff) by explicit guidance on applying fairness criteria to the review of materials for particular tests and content within tests. For example, reviewers examining Reading Test passages are informed that, in partial modification of the "Language" criterion above, a limited amount of foreign words and phrases, slang, dialect, and/or idiomatic expressions may be acceptable in Reading Test passages selected from works of literature and literary nonfiction, provided that sufficient context to enable understanding is available, because such elements are an authentic part of the real-world texts being sampled for the test.

While the general and specific criteria discussed above are intended to be the primary basis for the qualitative evaluation of test material for fairness, College Board encourages reviewers to draw on their professional judgment and expertise in order to apply the criteria flexibly and contextually. Moreover, reviewers are invited to raise issues that may not fall neatly into any of the above categories as part of the effort to ensure that all potential fairness issues receive thoughtful consideration.

Each fairness review of ACCUPLACER materials is first conducted individually, followed by a discussion via meeting held remotely. Reviewers provide comments in advance of the meeting; these comments are read and considered by College Board staff, who prepare potential responses, such as edits or removal, for discussion at the meeting. College Board staff raise issues that were identified by reviewers as high priority, commented on by multiple reviewers, or not identified as high priority but represent a potentially serious issue; guided by College Board staff, reviewers talk over these issues, evaluate College Board–proposed remedies when warranted, and introduce issues of their own, either ones previously mentioned in advance comments or those newly discovered. Whether comments are made only in writing or are also shared in a meeting, College Board staff carefully assess all feedback, make decisions informed by best practices and expert consensus, and produce records of how particular issues were resolved.

Pretesting. All questions are pretested on a motivated sample of test takers that resembles the population of interest and is sufficient in size to allow College Board to evaluate the materials statistically in terms of difficulty, to discern whether the questions can differentiate between lower- and higher-achieving test takers, and to ensure that test takers from different racial/ethnic and gender groups don't differentially respond

to the questions. The questions are embedded into operational test administrations and administered to test takers. The data from at least 1,000 test takers responding to each question are used to evaluate the performance of the questions. Once questions have been pretested and the statistics associated with them have been computed, the materials are reviewed by measurement and content specialists for content soundness, fairness, statistical discrimination, difficulty, and differential performance among groups of tested students.

DIF analysis. Analyses of differential item functioning, or DIF, are conducted on the items at the pretest stage to identify questions that may function differently for members of different groups. The underlying assumption in conducting such analyses is that all test takers demonstrating the same level of achievement in the content area should have similar chances of answering each question correctly, regardless of gender or race/ethnicity. DIF occurs when individuals with similar scores on an assessment differ notably in their performance on a specific test question by student subgroup. The presence of DIF indicates that a question functions differently for individuals from one subgroup from the way it functions for those of another subgroup who are at the same individual achievement level. Items exhibiting high levels of DIF may be measuring factors irrelevant to an assessment (such as culture) or more than one dimension for which the two groups have different strengths.

DIF analyses begin by examining any differences in the performance on each individual question of two comparable achievement level groups, referred to as the reference group and the focal group. Questions having extreme values of DIF, or those questions appearing to favor one group over another for test takers of the same level of achievement, undergo further review to determine whether some aspect of what the question is measuring is particularly related to subgroup membership and irrelevant to the dimension being measured. When a question is identified as exhibiting such characteristics, it's either revised and re-pretested, or eliminated. For more information on DIF as it relates to the ACCUPLACER Suite of Assessments, see Chapter 3: Test Development Procedures.

Operational administration

Fairness also involves equality in test administration across all groups of test takers. Without such standardization, the accuracy and comparability of score interpretations would be reduced (AERA, APA, & NCME, 2014). To ensure consistency in the administration of ACCUPLACER despite the different ways a test within the Suite may be administered—as a computer-adaptive test or a linear, accommodated (COMPANION) form; at the institution where a student is enrolled or one that's convenient and authorized to deliver the test; with a human proctor or a virtual one—College Board provides guidelines to ensure standardized administration in publicly available manuals and guides: the ACCUPLACER COMPANION Administrator's Manual (College Board, 2017a), ACCUPLACER User's Guide (College Board, 2017b), and ACCUPLACER Program Manual (College Board, 2018).

Standard 3.4 (AERA, APA, & NCME, 2014) states that issues of test security can also threaten the comparability of treatment of individuals and the validity and fairness of test score interpretations. College Board has put in place security measures—including the manuals and guides mentioned above, with detailed security guidelines and processes—to ensure that no test taker or group of test takers obtains access to information or opportunities that allow them to attain scores by fraudulent means and jeopardize the validity of the results of the assessment.

In September 2018 College Board launched the ACCUPLACER Certificate of Test Administration (ACTA) assessment to further help standardize the operational administration and security of ACCUPLACER. This mandatory assessment measures testing personnel's eligibility and readiness for administering ACCUPLACER tests by assessing their knowledge of standard ACCUPLACER processes and procedures, features of the ACCUPLACER platform, and steps for ensuring test security.

Predictive validity analyses for ACCUPLACER

Fairness extends beyond question/task performance and test construction and is strongly tied to validity. Standard 3.7 (AERA, APA, & NCME, 2014) addresses the notion that fair assessments ensure validity of test score interpretations as a basis for predicting future performance. In the case of ACCUPLACER, test scores should not provide different criterion prediction for different subgroups. It is then the responsibility of College Board to investigate the possibility of differential prediction. To this end, the plan for predictive analysis for ACCUPLACER tests includes exploratory investigation of differential predictions. This will be part of the collection of evidence based on criterion-related information mentioned in Chapter 7.

Score reporting and interpretation

A critical aspect of fairness is the fair and valid interpretations of test scores for intended uses (AREA/APA/NCME 2014, p. 53). To support appropriate interpretation and inferences made on the basis of ACCUPLACER test scores, College Board provides detailed score reports to test takers (in the form of individual score reports) and test users (in the form of institutional roster reports). Both types of reports are described in this section.

Along with publicly available interpretative materials developed by College Board, including content specifications, Skills Insight, Proficiency Statements, writing score and dimension descriptions (described in this chapter), and intended uses of ACCUPLACER and ACCUPLACER test scores (presented in Chapter 1), score reports provide crucial information on students' preparedness for postsecondary work. Collectively, they are designed to make sure all students have access to the information they require so they may seek the appropriate just-in-time academic supports they need to connect them to success in college and career training programs, and to help instructional programs in efforts to make such supports available. By providing such critical information to all test takers and the institutions and programs that serve them, College Board makes available to every test taker an equal opportunity to achieve success in college and in workforce training programs.

Score reporting

In keeping with the AERA/APA/NCME Standards, ACCUPLACER score reports have been developed at the student and institutional levels to provide their intended audiences with appropriate interpretations of the reports and guidelines outlining the appropriate use of test results. A variety of reports are available online 24/7 for all ACCUPLACER tests, including the following:

- Individual Score Report (ISR)
  - Generated for each student at the end of testing.
  - Shows student's identifying information and test scores, with CSEMs if option is selected by the institution.
  - Shows appropriate course placement, if placement rules have been entered into the testing site.
  - If WritePlacer is taken, the ISR includes the holistic score description and dimension statements.
- WritePlacer Response Report – Allows an institutional user to search and print essays submitted by students in response to a WritePlacer prompt.
- Placement Roster Report – Provides a list of students who placed into courses associated with a specific course group.
- Course Roster Report – Provides a list of students who placed into a specific course.
- Score Roster Report – Customizable report that may include students' scores, demographic information, and answers to background questionnaires as selected by the user for a specific date range.

Score Interpretation

College Board provides information to facilitate appropriate interpretations and understanding of scores received on ACCUPLACER tests.

Skills Insight

Skills Insight provides sets of data-driven statements intended to facilitate interpretation of performance on the five ACCUPLACER placement tests: Reading, Writing, Arithmetic, QAS, and AAF. The statements are organized by the following test score ranges, or bands:

- 236 and below
- 237–249
- 250–262
- 263–275
- 276 and above

The insight provided allows educators and students to see the skills and knowledge typically mastered at each score band so they may develop appropriate strategies for improvement. To look at the Skills Insight for the five ACCUPLACER placement tests, see Appendix C: Skills Insight. For more information on, including the development of, Skills Insight, see Chapter 5: Interpretation and Application of Results.

Proficiency Statements

Proficiency Statements provide sets of data-driven statements intended to facilitate interpretation of performance on the four multiple-choice ACCUPLACER ESL tests: Reading Skills, Sentence Meaning, Language Use, and Listening. Proficiency Statements were derived by convening a panel of experts in each subject area to review questions anchored at specific points along the scaled score range. The statements describe the knowledge and skills required to answer the anchored questions correctly. These statements offer useful information for understanding a test taker's achievement levels. However, as mentioned earlier, actual placement decisions should include other variables that may contribute to an accurate assessment of a student's achievement, such as high school grades and background information.

These statements allow educators and students to see the skills typically mastered at each score band so they may develop appropriate strategies for improvement. For a look at the Proficiency Statements for the ACCUPLACER ESL tests, see Appendix D: ACCUPLACER ESL Tests - Proficiency Statements

Writing score and dimension descriptions

WritePlacer and WritePlacer ESL provide descriptions of holistic scores received by students. In these, each writing sample is evaluated based on its overall impression and effectiveness evidenced by how well it communicates a whole message, not on the basis of the individual writing characteristics in isolation. An institution may also choose to have the more detailed writing dimension scores and descriptions included in the ISR.

These detailed descriptions allow educators and their students to see the writing skills typically observed in essays at each score point so that they may develop appropriate strategies for improvement. For an in-depth look at these descriptions, see Appendix A: Scoring Rubrics—WritePlacer and WritePlacer ESL.

Item challenge process

The ACCUPLACER Suite's item challenge process offers additional transparency and a check on the soundness and fairness of test materials. Test takers may alert proctors to potential issues with test materials. These are then forwarded to College Board through established email channels. Such queries are routed to senior test development staff, who review the materials in question and, if applicable, develop answer explanations. If the question is from a study or practice resource, the reporting test administrator is supplied with the review's findings; if the question is still being used on an operational

test, the reporting personnel is advised that the material was reviewed and what the outcome of the process was. In the rare circumstance in which a review identified a problem with the materials, College Board would undertake appropriate remediating action, up to and including removing problematic material from an item pool.

Practice

A critical aspect of test fairness rooted in but extending beyond design, development, and administration is practice. "Practice" includes the vital area of test familiarization—that is, making students aware of and comfortable with test instructions, formats, delivery methods, and the like. Resources and activities also focus on the underlying skills and knowledge fundamental to test constructs and hew closely to the test itself, their main purpose being to prepare test takers for the material they will encounter on test day.

Ensuring that all test takers have access to accurate, thorough information about the test well in advance of test day helps foster the goal of equity by giving everyone an equal chance to learn what is expected of them on the assessment, to address skill and knowledge gaps well in advance of the assessment, and to avoid wasting valuable test time on reading directions, figuring out what the questions are asking, and trying to understand how to navigate computer-delivered tests.

Test makers have a clear ethical and practical obligation to acquaint test takers and other stakeholders with their testing instruments. To this end, College Board provides a wealth of practice-related resources for tests in the ACCUPLACER Suite, all of them free of charge, to students and other stakeholders. These include downloadable sample test questions, including answer explanations, that are reviewed and approved by College Board (free online PDFs); practice questions delivered via a Study App, including answer explanations (free and computer based); information for interpreting test scores in the form of Skills Insight and Proficiency Statements after a student has tested; full test specifications for the ACCUPLACER placement tests; user manuals; and program manuals (free online PDFs). Additionally, College Board offers free professional development webinars designed to support faculty and test administrators who work with ACCUPLACER test takers. For more information on the study materials and information provided, see Section 1.2.

Accessibility

College Board is strongly committed to the concept of making its ACCUPLACER Suite test materials maximally accessible to all test takers. The organization subscribes to the principles of universal design, which, as noted by the AERA/APA/NCME Standards, has as its goal "[developing] tests that are as usable as possible for all test takers in the intended population, regardless of characteristics such as gender, age, language background, culture, socioeconomic status, or disability" (AREA/APA/NCME 2014, p. 57). To make sure test takers with disabilities can access the tests with the support needed, all ACCUPLACER materials are created in a highly legible layout. Other additional steps taken include:

- **The availability of assistive devices for online testing.** Accessibility Wizard, which makes it possible for test takers with visual impairments to change the appearance of testing screens for easier viewing, is available in the online test environment. It gives test takers the ability to select a display that enhances the legibility of test materials presented, including the ability to choose a high-contrast color scheme, font and cursor/point size and color, and line spacing. Institutions that use other assistive devices as a standard accommodation for students whose vision impairments prevent them from accessing screen content or navigating with a mouse may elect to use these for administering ACCUPLACER. Apart from providing the Accessibility

Wizard, the ACCUPLACER platform incorporates Read&Write Gold,[3] the NonVisual Desktop Access (NVDA) Screen Reader,[4] ZoomText® Magnifier/Reader,[5] Kurzweil 3000,[6] and JAWS® (Job Access With Speech).[7]

- **ADA compliance.** All print materials, including COMPANION tests, ACCUPLACER guides and manuals, sample questions/tasks, and test specifications are ADA compliant and can be used either in print forms or, if viewed on the computer, accessed through screen reader software.

As accessibility remains an ongoing focus for College Board, refinements to the print and digital presentation and delivery of test materials will continue to be made as College Board finds ways of making the tests more accessible to all.

Accommodations

While observing the principles and adopting the practices of universal design and of accessibility more generally are helpful in reducing the number and severity of construct-irrelevant barriers for all test takers, some test takers may still need additional support in order to complete the assessment and/or obtain valid test scores. Standards 3.9 through 3.14 discuss the responsibility of test developers to develop and provide test accommodations and the appropriate use of said accommodations (AERA, APA, & NCME, 2014, p. 67–70).

To provide a fair testing environment for all test takers, students with disabilities that affect their ability to participate in the ACCUPLACER Suite assessments are eligible to test with the accommodations they need. While College Board takes steps to ensure ACCUPLACER is accessible to all test takers, testing programs at individual institutions, acting on the counsel of their Services for Students with Disabilities (SSD) coordinators and taking students' documented needs into consideration, are charged with determining and offering test takers with accommodations for testing, such as large print, braille, audio recording, or a human reader, signer, or scribe. These testing programs, with support from College Board as needed, are entrusted with the responsibility to conduct testing without changing the construct or constructs being measured, such that scores maintain their meaning across all subgroups as well as for both accommodated and non-accommodated students. This practice ensures that, when appropriate and possible, construct-irrelevant barriers that can interfere with test takers accurately demonstrating their true standing on a construct are removed (AERA, APA, & NCME, 2014). As previously discussed, a construct-irrelevant barrier is any factor unrelated to the concepts or characteristics the assessment is designed to measure that can lead to an unfair testing experience and distort test takers' scores, decreasing the validity of the scores for their intended uses.

The accommodations offered by each institution's testing program serve to remove unfair disadvantages for those students with disabilities who have documented needs, allowing them to use accommodations on College Board assessments. In keeping with the AERA/APA/NCME Standards and best practices, accommodations are intended to "respond to specific individual characteristics, but [do] so in a way that does not change the construct the test is measuring or the meaning of scores" (AERA, APA, & NCME 2014, p. 67). To this end, all accommodated test formats 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 construct being tested and the meaning of the score remain unchanged.

The following are examples of accommodations intended to ensure eligible students receive the support they need. Please note that this list isn't exhaustive.

---

3. https://www.texthelp.com/en-us/products/read-write/
4. http://www.nvaccess.org/
5. http://www.aisquared.com/Products/index.cfm
6. http://www.kurzweiledu.com/kurz3000.aspx
7. http://www.freedomscientific.com/Products/Blindness/Jaws

Presentation

- Accommodated tests called ACCUPLACER COMPANION forms,[8] developed by College Board, are available for test takers who aren't able to take the computer-adaptive tests. They're available as linear tests in the following forms:
  - PDFs (retrievable off the secure ACCUPLACER platform)
  - Digital (taken on the computer with approved screen readers)
  - Regular print
  - Large print
  - Braille
  - Audio recording (CD)
  - Reader and reader script (Note: Reader reads entire test)

Many accommodations are administered in the standard testing room, as part of the administration. When judged appropriate, an SSD coordinator may administer accommodated tests in a separate setting.

Setting

- Small-group setting
- Private room
- Alternative testing site (with a human proctor, either in person or virtual)
- Preferential seating

Responding

- Verbal; essay dictated to scribe
- Answer sheet for computer-adaptive multiple-choice tests

Timing and Scheduling

- Frequent breaks
- Specified time of day

In general, students approved by SSD to receive testing accommodations meet the following criteria:

- **Student has a documented disability.** Examples of disabilities include, but aren't limited to, visual impairments, learning disorders, physical and medical impairments, and motor impairments. A student must have documentation of their disability, such as a current psychoeducational evaluation or a report from a doctor. The type of documentation needed depends on the student's disability and the accommodations being requested.

- **Participation in an assessment is impacted.** The disability must result in a relevant functional limitation that impacts the student's ability to participate in an ACCUPLACER assessment. For example, a student whose disabilities preclude sitting for extended periods may need accommodations, given that several tests are typically administered consecutively in a single sitting.

- **Requested accommodation is needed.** The student must demonstrate the need for the specific accommodation requested. For example, a student requesting to be tested alone in a private room should have documentation showing that they have difficulty performing test tasks in an open setting.

2.4 Conclusion

As the above discussion attests to, test fairness is a multidimensional concept that has important implications for nearly every aspect of test design, development, administration, and reporting. The concept of fairness includes but extends far beyond requirements to ensure that individual test questions are fair for all test takers. It also

---

8. Special accommodations available for test takers taking the test on the computer include assistive technology (screen reader), visual magnification, and the ability to change the size of fonts and backgrounds.

obligates test makers to define what they are measuring, for whom, and for what uses; to undertake exacting test design and development processes that promote the validity and fairness of all test materials; to identify subgroup performance differences and establish that test scores have the same meaning for members of different subgroups; to ensure that all students, including those with special needs, have the fullest possible access to the assessments; and to provide supports for all students so that they are as prepared as possible for test day, know and can interpret their scores, and have a clear sense of how to improve their performance as well as develop the underlying skills and knowledge being measured.

College Board has an abiding commitment to test fairness in all its many forms for the ACCUPLACER Suite of Assessments. In accordance with the AERA/APA/NCME Standards and best practices more generally, and with a keen sense of its profound ethical obligation to the students in its care and to the institutions and individuals who rely on and trust its data, College Board has undertaken numerous steps in the design of the ACCUPLACER Suite assessments to achieve the goal of test fairness. As new needs arise and as better approaches are discovered, the organization will continue to evolve, ensuring that its approaches meet or exceed the highest standards in the field.

Test Development Procedures

College Board seeks to make the ACCUPLACER Suite of Assessments deeply reflect the work that students need to do to be ready for and to succeed in college and their career paths. The individual questions and tasks and the tests as a whole reinforce valuable schoolwork, reflect a deep commitment to craft, and help define the level of rigor required for students to be college and career ready.

College Board works with various committees throughout the test design and development process to ensure the highest-quality assessments. Academic advisory committees and test development committees, which include secondary and postsecondary classroom teachers, provide input on test designs and test specifications, and help define the academic preparation needed for college. External content experts and fairness review panels perform independent review of each question before it is pretested and considered for operational use. When reviewing test questions, independent reviewers help to ensure that these are measuring important skills and knowledge, that the skills and knowledge the questions are assessing align well with the test specifications in terms of content and rigor, that all test questions are fair to all students, and that the questions are written in a way that model good instruction for the teacher and productive practice for the student.

College Board works with high school teachers and postsecondary instructors who teach entry-level, credit-bearing courses to develop appropriately challenging questions for the test that represent the kinds of tasks that students need to be able to perform if they're to be successful in higher education. Using items reviewed by these educational professionals helps to ensure that the items are relevant to the work students do in challenging classrooms and reflect the kinds of tasks that students will be expected to perform in college and workforce training.

This chapter details the steps in the ACCUPLACER Test Development process, from the establishment of test specifications until the tests are ready to be administered to test takers for a score. Section 3.1 details the guiding principles upon which this test development process is based. Section 3.2 discusses the test specifications for each test in the ACCUPLACER Suite. Section 3.3 details the creation, review, and analysis of the questions that make up the tests. Section 3.4 addresses the ACCUPLACER computer-adaptive algorithm that determines how those items are selected for administration to test takers. Finally, Section 3.5 details special accommodations for test takers who require alternative formats, including the development of accommodated forms called COMPANION, and conversion tables to facilitate comparison of scores from computer-adaptive and COMPANION tests.

3.1. Guiding Principles of College Board's Test Development Process

To achieve the vision outlined earlier in this chapter, every test in the ACCUPLACER Suite of Assessments is developed with care and expertise at every stage of the process. To that end, College Board has implemented a test development process that helps ensure that our ACCUPLACER Suite of Assessments questions and tasks are:

- Evidence-based and focused on the core set of knowledge and skills that are most important to prepare students for the rigors of college and career;
- Measuring student skills and knowledge as directly and authentically as possible by employing a range of question and task types relevant to instruction and life;

- Worth doing, crafted out of rich and engaging passages and contexts, reflective of best instructional practices, and rewarding of the academic excellence that any student can attain through deliberate practice;

- As motivating, interesting, engaging, and relevant to students as possible;

- Written by experts, many of whom have teaching experience at the middle school, high school, and postsecondary levels;

- Reviewed by multiple independent experts active in the field of education for content and fairness issues prior to pretesting;

- Accessible and fair to all students, having been developed to be content relevant, accurate, authentic and respectful in representation, and consistent with universal design principles.

The test development process for the ACCUPLACER Suite of Assessments is based on these guiding principles.

3.2. Test Specifications
ACCUPLACER Placement Tests

Content Specifications for the ACCUPLACER Reading Test
Overall Claim for the Test

The ACCUPLACER Reading Test is designed to collect evidence in support of a broad claim about student performance:

Students can demonstrate college and career readiness proficiency in reading and comprehending a broad range of appropriately challenging literary and informational texts in the content areas of literature and literary nonfiction, careers/history/social studies, humanities, and science.

Test Description

The Reading Test is primarily composed of passages with associated questions and a small number of questions that test vocabulary in sentence-length contexts. To answer the questions, test takers must refer to what the passages say explicitly and use careful reasoning to draw supportable inferences from the passages.

The texts used in reading passages are either commissioned (texts written specifically for the test) or authentic (texts excerpted or adapted from previously published sources). They may be one of two genres: literary or informational. Literary passages take the form of prose fiction or literary nonfiction (e.g., selections from memoirs and personal essays).[9] The informational genre includes texts spread across a defined range of content areas: careers/history/social studies (e.g., texts about career-related topics as well as texts in the academic social science disciplines), humanities (e.g., texts about arts and letters), and science (e.g., texts in the academic natural science disciplines). Selected texts represent one of three writing modes: narrative passages use chronology or sequence as the organizing principle; Informative/explanatory passages convey or describe new information and ideas; and Argumentative passages seek to move readers to action or to change belief through logical argumentation.

Questions assess test takers' understanding of information and ideas in these readings and measure key reading competencies deemed most relevant to future postsecondary success. They're classified into four content categories: Information and Ideas questions focus on the informational content of passages; Rhetoric questions focus on craft, structure, and authors' techniques, and may require test takers to analyze how word or phrase choices influence meaning, shape mode and tone, reflect point of view, or lend precision; Synthesis questions focus on the analysis of information and ideas between

---

9. Poetry and drama aren't represented.

multiple, topically related passages; and Vocabulary questions focus on the meaning of words and phrases as they're used in context. Answers are derived from what is stated or implied in the contexts rather than on prior knowledge of the topics.

In conclusion, the Reading Test is a challenging, carefully constructed assessment of comprehension and reasoning skill with an unmistakable focus on close reading of appropriately challenging passages in a wide array of subject areas.

Visit the ACCUPLACER practice website to see sample questions: **https://accuplacer.collegeboard.org/student/practice**.

Test Summary

The following tables (Table 3.1 and 3.2) provide a synopsis of key content dimensions of the Reading Test.

Table 3.1: *ACCUPLACER Reading Test Content Specifications*

Content Areas	Number of Questions	Percentage of Test*
Set-based questions		
1 literary set	4	20
1 informational paired set	4	20
Discrete questions		
Informational	12	60
Total	20	100
Question content distribution		
Information and Ideas	7–11	35–55
Rhetoric	7–11	35–55
Synthesis	2	10
Vocabulary	2–4	10–20

* Percentages don't necessarily add up to 100.

Table 3.2: *ACCUPLACER Reading Test Question Content*

Content Dimension	Description
Information and Ideas	
These questions focus on the informational content of text.	
Reading closely	The student will identify information and ideas explicitly stated in text and will draw reasonable inferences and logical conclusions from text.
Determining central ideas and themes	The student will identify explicitly stated central ideas and themes in text and determine implicit central ideas and themes from text.
Summarizing	The student will identify a reasonable summary of a text.
Understanding relationships	The student will identify explicitly stated relationships or determine implicit relationships between and among individuals, events, or ideas (e.g., cause-effect, comparison-contrast, sequence).
Rhetoric	
These questions focus on the craft and structure of writing.	
Analyzing word choice	The student will determine how the selection of specific words and phrases or the use of patterns of words and phrases shapes meaning and tone in text.
Analyzing text structure	The student will describe the overall structure of a text or analyze the relationship between a particular part of a text (e.g., a sentence) and the whole text.

(Continued)

Content Dimension	Description
Analyzing point of view	The student will determine the point of view or perspective from which a text is related or the influence this point of view or perspective has on content and style.
Analyzing purpose	The student will determine the main or most likely purpose of a text or of a particular part of a text (typically, one or more paragraphs).
Analyzing arguments	The student will analyze claims and counterclaims, assess an author's reasoning for soundness, and analyze how an author uses or fails to use evidence to support a claim or counterclaim.

Synthesis

These questions focus on synthesizing multiple sources of information.

Analyzing multiple texts	The student will synthesize information and ideas from multiple texts. (Note: All of the skills listed above may be tested with either single or paired passages.)

Vocabulary

These questions focus on determining the meaning of words and phrases in the contexts in which they appear.

Key Features of ACCUPLACER Reading Test

Distinctive features of the ACCUPLACER Reading Test are described below:

- Words in context
- Specified range of text length
- Specified range of text complexity
- Diversity

Words in context. Some questions on the Reading Test measure test takers' understanding of the meaning and use of words and phrases in the context of prose passages. These words and phrases are neither highly obscure nor specific to any one domain. Instead, they're words and phrases whose specific meaning and rhetorical purpose are derived in large part through the context in which they're used.

Text length. Reading passages fall into one of four length levels, as determined by a standard word count formula in which a "word" is defined as six characters. Very short passages are 75 to 100 standard words; short passages are 150 to 200 standard words; medium-length passages are 250 to 300 standard words; and long passages are 350 to 400 standard words.

Text complexity. Reading passages exhibit a defined range of text complexity from early high school level to postsecondary entry. To ensure that texts are appropriately challenging, test development staff use quantitative and qualitative measures of text complexity as well as feedback from secondary and postsecondary subject matter experts and test data on student performance. The computer-adaptive test design, to some extent, influences the distribution of text complexity encountered by any given test taker. The qualitative text complexity rubric can be found in Appendix B: Text Complexity (Qualitative)—Reading and Writing.

Diversity: College Board is committed to presenting students with a test-taking experience that is reflective of the diversity of the United States and the world. To that end, passage and question pools include substantial content that visibly reflects U.S.-based racial and ethnic diversity (including African American/black, American Indian/Native American, Asian American, and Latino/a individuals, culture, and experiences); international or global (non-U.S.) perspectives, cultures, or settings; and balanced representation of genders.

Content Specifications for the ACCUPLACER Writing Test

Overall Claim for the Test

The ACCUPLACER Writing Test is intended to collect evidence in support of a broad claim about student performance:

> *Students can demonstrate college and career readiness proficiency in revising and editing a range of texts in a variety of content areas, both academic and career related, for expression of ideas and for conformity to the conventions of standard written English grammar, usage, and punctuation.*

Test Description

National curriculum surveys conducted by College Board and others demonstrate that postsecondary instructors rate high in importance a student's ability to analyze an entire text and evaluate it in context. In accordance with this, the Writing Test is designed to comprise a series of passages, and test takers are required to answer questions based on extended-prose contexts rather than in isolation or in limited (e.g., single-sentence) contexts. Although some questions are answerable by referring to a single phrase, clause, or sentence, many others leverage the extended context the test's format makes available and require students to have an understanding of multiple sentences, one or more paragraphs, or the passage as a whole.

All passages on the Writing Test are written specifically for the test so that "errors" (a collective term for various content-related, rhetorical, or mechanical problems) can be introduced that students must recognize and correct. In their base, "correct" form, Writing passages are well-written essayistic prose pieces on topics in literary nonfiction, careers/history/social studies, the humanities, and science, with the core writing modes of argument, informative/explanatory text, and nonfiction narrative represented.

Questions associated with these passages measure key writing competences deemed most relevant for future postsecondary success by placing test takers in the role of someone who must make authentic, context-based revision and editing decisions to the work of an unspecified writer. Test takers are, by turns, asked to improve the development, organization, and use of language in the passages and to ensure that the passages conform to conventions of standard written English grammar, usage, and punctuation. For example, questions may ask test takers to establish or refine the central points of passages; add, revise, or delete supporting material; and improve the focus of passages. Some questions ask test takers to enhance the rhetorical effectiveness of language by making contextually based choices that improve precision and concision; maintain consistency in style and tone or achieve particular stylistic effects; and combine sentences and rearrange sentence elements in ways that improve clarity and accomplish specified rhetorical goals, such as placing emphasis on a main point rather than a subordinate point.

Answers to all questions are anchored in the context of the passages. Neither rote recall of language rules nor context-free applications of grammar, usage, and mechanics conventions are tested.

Visit the ACCUPLACER practice website to see sample questions: **https://accuplacer.collegeboard.org/student/practice**.

Test Summary

Writing Test questions are classified into two clusters, each of which includes three content categories.

The Expression of Ideas cluster contains Development questions (which address revision of text to improve structure, support, and focus); Organization questions (which address revision of text to improve logical sequencing as well as achieving effective introductions, conclusions, and transitions); and Effective Language Use questions (which address

revision of text to improve the precision and concision of expression, to maintain style and tone, or to meet particular rhetorical goals, and to use syntax to achieve specified rhetorical purposes).

The Standard English Conventions cluster contains Sentence Structure questions (which address editing of text to improve sentence formation and consistency in construction); Conventions of Usage questions (which address editing of text to make it conform to standard written English usage conventions); and Conventions of Punctuation questions (which address editing of text to make it conform to standard written English punctuation conventions).

The following tables (Tables 3.3 and 3.4) provide a synopsis of key content dimensions of the Writing Test.

Table 3.3: *ACCUPLACER Writing Test Content Specifications*

Content Areas	Number of Questions	Percentage of Test*
Set-based questions		
1 literary set	5	20
4 informational sets	20	80
Total	25	100
Question content distribution	14–16	56–64
Expression of Ideas		
▪ Development		
▪ Organization		
▪ Effective Language Use		
Standard English Conventions	9–11	36–44
▪ Sentence Structure		
▪ Conventions of Usage		
▪ Conventions of Punctuation		

* Percentages don't necessarily add up to 100.

Table 3.4: *ACCUPLACER Writing Test Question Content*

Content Dimension	Description
Development	
These questions focus on revising text in relation to rhetorical purpose. (Prior knowledge of the topic is not assessed, though consistency of the material within a passage may be.)	
Proposition	The student will add, revise, or retain central ideas, main claims, topic sentences, and the like to structure texts and to convey arguments, information, and ideas clearly and effectively.
Support	The student will add, revise, or retain information and ideas (e.g., details, facts, statistics) intended to support claims or points in text.
Focus	The student will add, revise, retain, or delete information and ideas in text for the sake of relevance to topic and purpose.
Organization	
These questions focus on revision of text to improve the logic and cohesion of text at the sentence, paragraph, and whole-text level.	
Logical sequence	The student will revise text as needed to ensure that information and ideas are presented in the most logical order.

(Continued)

Content Dimension	Description
Introductions, conclusions, and transitions	The student will revise text as needed to improve the beginning or ending of a text or paragraph or to ensure that transition words, phrases, or sentences are used effectively to connect information and ideas.

Effective Language Use

These questions focus on revision of text to improve the use of language to accomplish particular rhetorical purposes.

Precision	The student will revise text as needed to improve the exactness or content appropriateness of word choice.
Concision	The student will revise text as needed to improve the economy of word choice (i.e., to eliminate wordiness and redundancy).
Style and tone	The student will revise text as needed to ensure the consistency of style and tone within a text or to improve the match of style and tone to purpose.
Syntax	The student will use various sentence structures to accomplish needed rhetorical purposes.

Sentence Structure

Sentence boundaries	The student will recognize and correct grammatically incomplete sentences (e.g., rhetorically inappropriate fragments and run-ons).
Subordination and coordination	The student will recognize and correct problems in coordination and subordination in sentences.
Parallel structure	The student will recognize and correct problems in parallel structure in sentences.
Modifier placement	The student will recognize and correct problems in modifier placement (e.g., misplaced or dangling modifiers).
Shifts in verb tense	The student will recognize and correct inappropriate shifts and make appropriate shifts in verb tense within and between sentences.
Shifts in verb voice and mood	The student will recognize and correct inappropriate shifts and make appropriate shifts in verb voice and mood within and between sentences.
Shifts in pronoun person and number	The student will recognize and correct inappropriate shifts and make appropriate shifts in pronoun person and number within and between sentences.

Conventions of Usage

Possessive determiners	The student will recognize and correct cases in which possessive determiners (*its, your, their*), contractions (*it's, you're, they're*), and adverbs (*there*) are confused with each other.
Noun agreement	The student will recognize and correct lack of agreement between nouns.
Pronoun clarity	The student will recognize and correct pronouns with unclear or ambiguous antecedents.
Pronoun-antecedent agreement	The student will recognize and correct lack of agreement between pronoun and antecedent.
Subject-verb agreement	The student will recognize and correct lack of agreement between subject and verb.
Frequently confused words	The student will recognize and correct instances in which a word or phrase is confused with another (e.g., *accept/except*, *allusion/illusion*).

(Continued)

Content Dimension	Description
Logical comparison	The student will recognize and correct cases in which unlike terms are compared.
Conventional expression	The student will recognize and correct cases in which a given expression is inconsistent with standard written English.
Conventions of Punctuation	
End-of-sentence punctuation	The student will recognize and correct inappropriate uses of ending punctuation in cases in which the context makes the intent clear.
Within-sentence punctuation	The student will correctly use colons, semicolons, and dashes to indicate sharp breaks in thought within sentences.
Possessive nouns and pronouns	The student will distinguish among singular, singular possessive, plural, and plural possessive nouns and pronouns.
Items in a series	The student will correctly use and recognize and correct inappropriate uses of punctuation (commas and sometimes semicolons) to separate items in a series.
Nonrestrictive and parenthetical elements	The student will correctly use punctuation (commas, parentheses, dashes) to set off nonrestrictive and parenthetical sentence elements as well as recognize and correct cases in which restrictive or essential sentence elements are inappropriately set off with punctuation.
Unnecessary punctuation	The student will recognize and correct cases in which unnecessary punctuation appears in a sentence.

Key Features of ACCUPLACER Writing Test

Distinctive features of the ACCUPLACER Writing Test are described below:

- Words in context
- Command of evidence
- Specified range of text length
- Specified range of text complexity
- Diversity

Words in context. Some questions on the Writing Test measure the ability of test takers to apply knowledge of words, phrases, and language in general in the context of extended prose passages. Test takers may be asked to demonstrate they're able to use words carefully and with purpose (e.g., to improve precision and concision).

Command of evidence. The Writing Test measures test takers' capacity to revise a text to improve its development of information and ideas. To answer these questions, test takers must have a solid grasp of the content of the passage in question (although it's important to note that prior knowledge of the topic isn't expected).

Text length. Writing passages are between 300 and 350 standard words (based on a standard word count formula in which a "word" is defined as six characters), and typically contain 12 to 16 sentences.

Text complexity. Writing passages exhibit a defined range of text complexity from early high school level to postsecondary entry, with a large proportion of the passages reflective of college and career readiness level. To ensure that texts are appropriately challenging, test development staff use quantitative and qualitative measures of text complexity as well as feedback from secondary and postsecondary subject-matter experts and test data on student performance. The qualitative text complexity rubric can be found in Appendix B: Text Complexity (Qualitative)—Reading and Writing.

Diversity: College Board is committed to presenting students with a test-taking experience that is reflective of the diversity of the United States and the world. To that end, passage and question pools include substantial content that visibly reflects U.S.-based racial and ethnic diversity (including African American/black, American Indian/Native American, Asian American, and Latino/a individuals, culture, and experiences); international or global (non-U.S.) perspectives, cultures, or settings; and balanced representation of genders.

Content Specifications for the ACCUPLACER Arithmetic Test

Overall Claim for the Test

The ACCUPLACER Arithmetic Test is intended to collect evidence in support of the following claim about student performance:

> *Students can demonstrate fluency with, understanding of, and the ability to apply fundamental arithmetic concepts, skills, and practices that show progress toward success in college courses, career training, and career opportunities.*

Test Description

Questions on the Arithmetic Test focus on computation, order of operations, estimation and rounding, comparing and ordering values in different formats, and recognizing equivalent values across formats. The Arithmetic Test measures the students' knowledge and skills in these areas with a balance across fluency, conceptual understanding, and application. Application questions present a real-world context that requires students to demonstrate the ability to analyze a situation, determine the essential elements required to solve the problem, represent the problem mathematically, and carry out a solution. All questions are multiple-choice in format. The following knowledge and skill categories are assessed: whole number operations; fraction operations; decimal operations; percent; and number comparisons and equivalents.

Some questions on the Arithmetic Test allow the use of a basic four-function calculator, and others don't provide the calculator option. If a question is configured to allow for the use of a four-function calculator, the dropdown calculator icon will present in the top right corner of the screen. Students with a documented disability may use a handheld calculator if it's a prescribed accommodation.

Visit the ACCUPLACER practice website to see sample questions:
https://accuplacer.collegeboard.org/student/practice.

Test Summary

The following tables (Tables 3.5 and 3.6) provide a synopsis of key content dimensions of the Arithmetic test.

Table 3.5: *ACCUPLACER Arithmetic Test Content Specifications*

Content Areas	Number of Questions	Percentage of Test
Whole number operations	3–5	15–25
Fraction operations	3–5	15–25
Decimal operations	3–5	15–25
Percent	3–5	15–25
Number comparisons and equivalents	3–5	15–25
Total	20	100

Table 3.6: *ACCUPLACER Arithmetic Test Question Content*

Content Dimensions and Descriptions

Whole number operations

Addition, subtraction, multiplication, and division of whole numbers, including order of operations, estimation and rounding, and applying operations to real-life contexts

Fraction operations

Addition, subtraction, multiplication, and division of fractions and mixed numbers, including order of operations, estimation and rounding, and applying operations to real-life contexts

Decimal operations

Addition, subtraction, multiplication, and division of decimal numbers, including order of operations, estimation and rounding, and applying operations to real-life contexts

Percent

Calculation with percent with or without a context, including percent increase, percent decrease, determining the percent of a number, and applying percent to real-life contexts

Number comparisons and equivalents

Comparisons of differently formatted values by ordering, using the number line, and using equality/inequality symbol notation; and evaluation of equivalent number statements (to assess mental math strategies)

Key Features of ACCUPLACER Arithmetic Test

Distinctive features of the Arithmetic Test are described below:

- Math that matters most
- Calculator and no-calculator questions
- Problems grounded in real-world contexts
- Specified range of text length

Math that matters most. Questions on the Arithmetic Test focus on skills that are essential for the path to college and career readiness. These include an emphasis on applied reasoning questions over reasoning questions disconnected from the math curriculum as well as a strong emphasis on both fluency with mathematical procedures and conceptual understanding.

Calculator and no-calculator questions. The Arithmetic Test includes both questions that restrict the use of a calculator and questions that allow the use of a basic, four-function calculator. No-calculator questions assess fluency in rational number arithmetic and include conceptual questions for which a calculator isn't needed, while questions with a calculator option give insight into students' capacity for strategic use of the tool to address problems efficiently.[10]

Real-world contexts. Questions in the Arithmetic Test include a proportion of contextualized questions needed for allowing test takers to engage with questions related to work performed in college and career, and to mitigate the disconnection between math concepts and real-life applications for students.

Text length. Care is taken to ensure that contextualized questions measure the relevant math construct only, and that the questions do not measure reading skills or abilities. First, the test design specifies three word-count levels: Low (fewer than 40 words; these include questions that are not contextualized), Medium (40–60 words), and High (more than 60 words). The question pool is developed to ensure that the majority of questions have Low or Medium word count. Finally, constraints are implemented in the item-selection algorithm so that questions with Medium and High word count are capped. On the 20-question Arithmetic test, a test taker may receive up to 20 (i.e., all) questions coded Low for word count; questions coded Medium are capped at 5; and only 2 questions coded High could be administered.

---

10. While the use of handheld calculators shouldn't be allowed on any computer-adaptive ACCUPLACER Math Tests, test takers with a documented disability may use one if it's a prescribed accommodation.

Content Specifications for the ACCUPLACER Quantitative Reasoning, Algebra, and Statistics (QAS) Test

Overall Claim for the Test

The ACCUPLACER QAS Test is intended to collect evidence in support of the following claim about student performance:

> *Students can demonstrate fluency with, understanding of, and the ability to apply the mathematical concepts, skills, and practices that are most strongly prerequisite and central to their ability to progress through a range of college courses, career training, and career opportunities.*

Test Description

Questions on the QAS Test focus on a range of topics including computing with rational numbers, applying ratios and proportional reasoning, creating linear expressions and equations, graphing and applying linear equations, understanding probability and sets, and interpreting graphical displays. Test takers are required to exhibit command of mathematical practices, fluency with mathematical procedures, and conceptual understanding of mathematical ideas. The problems explore the full dynamic range of each content area through precisely crafted questions that emphasize the use of math in unlocking insights and solving problems. All questions are multiple-choice. In addition, questions may assess a student's math knowledge and skills via fluency with mathematical procedures and conceptual understanding. They may also present a real-world context that requires students to demonstrate the ability to analyze a situation, determine the essential elements required to solve the problem, represent the problem mathematically, and carry out a solution. All questions are multiple-choice in format. The following knowledge and skill categories are assessed: Rational numbers; Ratio and proportional relationships; Exponents; Algebraic expressions; Linear equations; Linear applications and graphs; Probability and sets; Descriptive statistics; Geometry concepts for Pre-Algebra; Geometry concepts for Algebra 1.

Some questions on the QAS Test allow the use of a calculator, and others don't have the calculator option. If a question is configured to allow for the use of a calculator, the calculator icon will present in the top right corner of the screen. For questions that are configured for multiple calculators, clicking on the icon will provide the student with a drop-down menu that could include two or three of the following: Basic (or four-function) calculator; square root calculator (or four-function calculator with a square root button); and TI-84 graphing calculator.

Visit the ACCUPLACER practice website to see sample questions:
https://accuplacer.collegeboard.org/student/practice.

Test Summary

The following tables (Tables 3.7 and 3.8) provide a synopsis of key content dimensions of the QAS Test.

Table 3.7: *ACCUPLACER Quantitative Reasoning, Algebra, and Statistics Test Content Specifications*

Content Areas	Number of Questions	Percentage of Test*
Rational numbers	1–3	5–15
Ratio and proportional relationships	3–4	15–20
Exponents	2–3	10–15
Algebraic expressions	2–3	10–15
Linear equations	2–4	10–20
Linear applications and graphs	2–4	10–20

(Continued)

Content Areas	Number of Questions	Percentage of Test*
Probability and sets	1–3	5–15
Descriptive statistics	1–3	5–15
Geometry concepts for Pre-Algebra	1–2	5–10
Geometry concepts for Algebra 1	1–2	5–10
Total	20	100

* Percentages don't necessarily add up to 100.

Table 3.8: *ACCUPLACER Quantitative Reasoning, Algebra, and Statistics Test Question Content*

Content Dimensions and Descriptions

Rational numbers

Calculating and applying rational numbers (with or without a context), including usage of absolute value

Ratio and proportional relationships

Calculating with rates, ratios, and proportions (with or without a context), and using unit conversions

Exponents

Calculating with exponents, radicals, fractional exponents, and applying scientific notation

Algebraic expressions

Creating and evaluating expressions to represent situations, and using properties of operations to combine like terms and identify equivalent expressions

Linear equations

Creating linear equations in one or two variables, solving linear equations, simplifying linear equations and inequalities, and solving systems of two linear equations

Linear applications and graphs

Applying linear equations to real-life contexts, using elementary linear functions to describe relationships, and graphing linear equations in two variables, linear inequalities, parallel and perpendicular lines, and systems of equations

Probability and sets

Calculating probability (simple, compound, and conditional), and defining sample spaces and events using set notation

Descriptive statistics

Interpreting graphical displays of data (histograms, box plots, and scatterplots), describing shape and spread of a sample set, and calculating measures of center

Geometry concepts for Pre-Algebra

Determining area and perimeter, circle area and circumference, and volume of prisms

Geometry concepts for Algebra 1

Creating expressions for area, perimeter, and volume, using distance formula and Pythagorean theorem, and evaluating basic geometric transformations

Key Features of ACCUPLACER QAS Test

Five distinctive features of the QAS Test are described below:

- Math that matters most
- Calculator and no-calculator questions
- Problems grounded in real-world contexts

- Specified range of text length
- Connection to math pathways

Math that matters most. The QAS Test focuses on skills that are essential for college and career readiness. These include an emphasis on applied reasoning questions over reasoning questions disconnected from the math curriculum as well as a strong emphasis on both fluency with mathematical procedures and conceptual understanding.

Calculator use. The QAS Test includes questions with no calculator option as well as questions with one or more calculator options. No-calculator questions assess fluency in rational number arithmetic and include conceptual questions for which a calculator isn't needed, while questions with calculator options give insight into students' capacity for strategic use of the tool to address problems efficiently.[11]

Real-world contexts. Questions on the QAS Test include a proportion of contextualized questions needed for allowing test takers to engage with questions related to work performed in college and career, and to mitigate the disconnection between math concepts and real-life applications for students.

Text length. Care is taken to ensure that contextualized questions measure the relevant math construct only, and that the questions do not measure reading skills or abilities. First, the test design specifies three word-count levels: Low (fewer than 40 words; these include questions that are not contextualized), Medium (40–60 words), and High (more than 60 words). The question pool is developed to ensure that the majority of questions have Low or Medium word count. Finally, constraints are implemented in the item-selection algorithm so that questions with Medium and High word count are capped. On the 20-question QAS test, a test taker may receive up to 20 (i.e., all) questions with Low word count, while questions coded Medium are capped at 5, and only 2 questions coded High could be administered.

Connection to math pathways. Because the math prerequisites of different college majors and career paths can be vastly different, the ACCUPLACER QAS placement test delivers a stronger connection to the wide variety of course sequences in both STEM and non-STEM fields of study. Students who struggle with skills and abilities related to algebra and functions will still be successful in many entry-level college courses. Therefore, the QAS test provides opportunities for students to demonstrate pre-algebra skills and abilities such as ratio & proportion, algebraic expressions, geometry, and probability.

Content Specifications for the ACCUPLACER Advanced Algebra and Functions (AAF) Test

Overall Claim for the Test

The ACCUPLACER AAF Test is intended to collect evidence in support of the following claim about student performance:

> *Students can demonstrate fluency with, understanding of, and the ability to apply more advanced mathematical concepts, skills, and practices that are most strongly prerequisite and central to their ability to progress through a range of college courses, career training, and career opportunities.*

Test Description

Questions on the AAF Test focus on a range of topics including a variety of equations and functions, including linear, quadratic, rational, radical, polynomial, and exponential. Questions will also delve into some geometry and trigonometry concepts. In addition, questions may assess a student's math knowledge and skills via fluency with mathematical procedures and conceptual understanding. They may also present a real-

---

11. While the use of handheld calculators shouldn't be allowed on any computer-adaptive ACCUPLACER Math Tests, test takers with a documented disability may use one if it's a prescribed accommodation.

world context that requires students to demonstrate the ability to analyze a situation, determine the essential elements required to solve the problem, represent the problem mathematically, and implement a solution. Test takers are required to exhibit command of mathematical practices, fluency with mathematical procedures, and conceptual understanding of mathematical ideas. The problems explore the full dynamic range of each content area defined in the content specifications through precisely crafted questions that emphasize the use of math in unlocking insights and solving problems. All questions are multiple-choice.

Some questions on the AAF Test allow the use of a calculator, and others don't provide the calculator option. If a question is configured to allow for the use of a calculator, the calculator icon will present in the top right corner of the screen. For questions that are configured for multiple calculators, clicking on the icon will provide the student with a drop-down menu that will include two or three of the following: Basic (or four-function) calculator; square root calculator (or four-function calculator with a square root button); and TI-84 graphing calculator.

Visit the ACCUPLACER practice website to see sample questions: **https://accuplacer.collegeboard.org/student/practice**.

Test Summary

The following tables (Tables 3.9 and 3.10) provide a synopsis of key content dimensions of the AAF test.

Table 3.9: *ACCUPLACER Advanced Algebra and Functions Test Content Specifications*

Content Areas	Number of Questions	Percentage of Test*
Linear equations	2–3	10–15
Linear applications and graphs	2–3	10–15
Factoring	1–2	5–10
Quadratics	2–3	10–15
Functions	2–4	10–20
Radical and rational equations	1–3	5–15
Polynomial equations	1–3	5–15
Exponential and logarithmic equations	1–3	5–15
Geometry concepts for Algebra 1	1–2	5–10
Geometry concepts for Algebra 2	1–2	5–10
Trigonometry	1–3	5–15
Total	20	100

* Percentages don't necessarily add up to 100.

Table 3.10: *ACCUPLACER Advanced Algebra and Functions Test Content*

Content Dimensions and Descriptions

Linear equations

Creating linear equations in one or two variables, solving linear equations, simplifying linear equations and inequalities, and solving systems of two linear equations

Linear applications and graphs

Applying linear equations to real-life contexts, using elementary linear functions to describe relationships, and graphing linear equations in two variables, linear inequalities, parallel and perpendicular lines, and systems of equations

(Continued)

Content Dimensions and Descriptions

Factoring

Factoring methods applied to quadratics, cubics, and polynomials

Quadratics

Creating quadratic equations in one or two variables, solving quadratic equations (via factoring or using the quadratic equation), simplifying quadratic equations and inequalities, and solving systems that involve a quadratic equation

Functions

Creating functions using function notation, evaluating linear and quadratic functions, graphing functions, and interpreting functions within a context

Radical and rational equations

Creating radical and rational equations and functions in one variable, determining domain and range for radical and rational functions, graphing radical and rational functions, and simplifying radical and rational expressions and equations

Polynomial equations

Creating polynomial equations in one and two variables, solving polynomial equations, and graphing polynomial functions

Exponential and logarithmic equations

Creating exponential and logarithmic equations in one and two variables, solving exponential and logarithmic equations, graphing exponential and logarithmic functions, and interpreting exponential and logarithmic functions

Geometry concepts for Algebra 1

Creating expressions for area, perimeter, and volume, using distance formula and Pythagorean theorem, and evaluating dilations, rotations, translations, and reflections

Geometry concepts for Algebra 2

Determining volume of nonprism objects, using intersecting line theorems, using triangle similarity and congruency theorems, and using circle equations in the coordinate plane

Trigonometry

Solving trigonometric equations, using right triangle trigonometry including special triangles, evaluating equivalent trigonometric functions, graphing trigonometric relationships, determining arc length and radian measures, and using the law of sines and the law of cosines

Key Features of ACCUPLACER AAF Test

Five distinctive features of the AAF Test are described below:

- Math that matters most
- Calculator and no-calculator questions
- Problems grounded in real-world contexts
- Specified range of text length
- Connection to math pathways

Math that matters most. The AAF Test focuses on skills that are essential for college and career readiness. These include an emphasis on applied reasoning questions over reasoning questions disconnected from the math curriculum as well as a strong emphasis on both fluency with mathematical procedures and conceptual understanding.

Calculator use. The AAF Test includes questions with no calculator option as well as questions with one or more calculator options. No-calculator questions assess fluency in rational number arithmetic and include conceptual questions for which a calculator isn't

needed, while questions with calculator options give insight into students' capacity for strategic use of the tool to address problems efficiently.[12]

Real-world contexts. Questions in the AAF Test include a proportion of contextualized questions needed for allowing test takers to engage with questions related to work performed in college and career, and to mitigate the disconnection between math concepts and real-life applications for students.

Text length. Care is taken to ensure that contextualized questions measure the relevant math construct only, and that the questions do not measure reading skills or abilities. First, the test design specifies three word-count levels: Low (fewer than 40 words; these include questions that are not contextualized), Medium (40–60 words), and High (more than 60 words). The question pool is developed to ensure that the majority of questions have Low or Medium word count. Finally, constraints are implemented in the item-selection algorithm so that questions with Medium and High word count are capped. On the 20-question AAF test, a test taker may receive up to 20 (i.e., all) questions with Low word count, while questions coded Medium are capped at 5, and only 2 questions coded High could be administered.

ACCUPLACER English as a Second Language (ESL) Tests

The ACCUPLACER ESL tests were developed as a response to an expressed need for an assessment measuring the skills of ELLs whose language skills wouldn't allow them to achieve meaningful and actionable scores on the college placement ACCUPLACER tests. They're thus intended for use in placing college-bound ELLs into appropriate ESL courses, so that they may receive the instruction needed to support them on their path to college and career. Courses into which students could be placed on the basis of their ESL test scores include ESL, developmental reading, and developmental English.[13]

ACCUPLACER ESL tests include Reading Skills, Sentence Meaning, Language Use, and Listening. With the exception of WritePlacer ESL, which is an essay test, all ACCUPLACER ESL tests are computer-adaptive, 20-question multiple-choice assessments.

Content Specifications for the ACCUPLACER ESL Reading Skills Test

Overall Claim for the Test

The ESL Reading Skills test is intended to collect evidence in support of the following claim about student performance:

> *ELLs can demonstrate their developed ability to read and comprehend appropriately challenging short passages and narratives in a variety of content areas by referring to what the texts say explicitly and drawing reasonable inferences from the passages to compensate for gaps in their lexical and structural control of the language.*

Test Description

The ESL Reading Skills Test contains short passages of 50 or fewer words and moderate-length passages of 50 to 90 words. Passages cover a range of contexts: science, history/social science, arts, humanities, narratives, psychology/human relations, and "practical situations." Less complex passages on the test tend to be straightforward and follow a predictable pattern of presentation; they're typically short, reflect primarily language found in everyday contexts, and use sentences containing high-frequency words. More complex passages on the test may reflect the use of less familiar subjects, some linguistic formulations found in academic readings, and a broader range of vocabulary, genres, and rhetorical structures.

---

12. While the use of handheld calculators shouldn't be allowed on any computer-adaptive ACCUPLACER Math Tests, test takers with a documented disability may use one if it's a prescribed accommodation.
13. Placement into college-level courses is typically based on the non-ESL ACCUPLACER college placement tests (i.e., ACCUPLACER Reading, Writing, or WritePlacer).

Two types of questions are asked on the ESL Reading Skills Test. Literal comprehension questions make up half the test. These questions assess a test taker's developed ability to comprehend a text and include questions requiring the test taker to identify and locate explicitly stated information and understand paraphrased information. Some literal comprehension questions test understanding of pronoun references. The remaining questions assess test takers' ability to make reasonable inferences. These include questions that ask test takers to read passages closely and identify main ideas, distinguish facts from opinions, determine cause-effect relationships, and infer the author's purpose, tone, and point of view. Some inference questions may ask test takers to determine the meanings of words and phrases from the contexts in which they're used.

Visit the ACCUPLACER practice website to see sample questions:
https://accuplacer.collegeboard.org/student/practice.

Test Summary
Table 3.11 provides a synopsis of key content dimensions of the test.

Table 3.11: *ACCUPLACER ESL Reading Skills Test Content Specifications*

Content Areas	Number of Questions	Percentage of Test*
Skills assessed		
Literal Comprehension	6–14	30–70
Inference	6–14	30–70
Passages support either one or two questions	6–6	30–30
One-question passage	7–7	35-35
Two-question passage		
Content areas		
Science	2–5	10–25
History/Social Science	2–5	10–25
Arts/Humanities	2–5	10–25
Narratives	2–5	10–25
Psychology/Human Relations	2–5	10–25
Practical Situations	2–5	10–25
Total	20	100

* Percentages don't necessarily add up to 100 because some questions meet multiple content requirements.

Key Features of the ESL Reading Skills Test
Distinctive features of the ESL Reading Skills Test are described below.

Skills that matter most. The ESL Reading Skills Test focuses on two critical comprehension skills. Test takers are assessed on their developed ability to comprehend information that is explicitly stated in a given text and their ability to reasonably infer from contextual clues information not directly stated.

Accessible contexts. As a test for students representing a wide range of cultural and linguistic backgrounds, experiences, and lengths of exposure to U.S. culture and English-speaking environments, the ESL Reading Skills Test uses contexts carefully developed to be as accessible and free of elements that would impede access to the content as possible. Prior topic-specific knowledge, including knowledge of U.S. culture or norms, isn't tested.

Accessible language. While language isn't simplified when it's part of the construct being assessed, passages and questions used on the test are carefully developed to be as clear and unambiguous as possible. For example, lower-frequency words and less common syntactic structures aren't included unless they're part of the construct tested.

Diversity. As this test is taken by students of diverse backgrounds, the question pool includes substantial content that visibly reflects U.S.-based racial and ethnic diversity as well as international or global perspectives, cultures, and settings. Items also reflect a balanced representation of genders.

Content Specifications for the ACCUPLACER ESL Sentence Meaning Test
Overall Claim for the Test

The ESL Sentence Meaning Test is intended to collect evidence in support of the following claim about student performance:

> *ELLs can demonstrate their developed ability to read and comprehend sentences written in a variety of structures on a variety of topics, compensating for gaps in their lexical and structural control of the language by using contextual clues.*

Test Description
The ESL Sentence Meaning Test assesses the understanding of the meaning of words and phrases in one- or two-sentence contexts. Sentences used on the test are drawn from the content areas of natural science, history/social studies, arts/humanities, psychology/human relations, and practical situations (the last typically dealing with familiar, everyday contexts).

Some questions ask test takers to render a sentence meaningful or grammatically correct by selecting a word or phrase from given options to fill in a blank. Such questions may require test takers to demonstrate an understanding of word order, word meaning, parts of speech, precedents and antecedents, and relationships (e.g., compare/contrast, cause/effect, sequencing, time frame, and chronology of events). Other questions present a short text made up of one or two sentences and may test students' developed ability to understand the gist or a detail presented in the sentence. Such questions may also ask test takers to show they're able to identify paraphrased information.

Four content areas are measured on the Sentence Meaning Test. These are Particles, Phrasal Verbs, and Prepositions; Adverbs, Adjectives, and Connectives Sequences; Basic Nouns and Verbs; and Basic and Important Idioms.

Visit the ACCUPLACER practice website to see sample questions: **https://accuplacer.collegeboard.org/student/practice**.

Test Summary
Table 3.12 provides a synopsis of key content dimensions of the test.

Table 3.12: *ACCUPLACER ESL Sentence Meaning Test Content Specifications*

Content Areas	Number of Questions	Percentage of Test*
Question types		
Fill in the blank	7–13	35-65
4-option multiple-choice	7–13	35-65
Content Assessed		
Particles, phrasal verbs, prepositions	2–6	10–30
Adverbs, adjectives, connectives sequence	3–8	15–40
Basic nouns and verbs	8–13	40–65
Common idioms	1–3	5–15
Total	20	100

* Percentages don't necessarily add up to 100 because some questions meet multiple content requirements.

Key Features of the ESL Sentence Meaning Test

Distinctive features of the ESL Sentence Meaning Test are described below.

- **Skills that matter most.** The ESL Sentence Meaning Test focuses on the critical reading skills of using elements in sentences, including parts of speech, linguistic markers indicating relationships and tenses, and transitional words and phrases, to aid comprehension and create meaning.

- **Accessible contexts.** As a test for students representing a wide range of cultural and linguistic backgrounds, experiences, and lengths of exposure to U.S. culture and English-speaking environments, the ESL Sentence Meaning Test uses contexts carefully developed to be as accessible and free of elements that would impede access to the content as possible. Prior topic-specific knowledge, including knowledge of U.S. culture or norms, isn't tested.

- **Accessible language.** While language isn't simplified when it's part of the construct being assessed, passages and questions used on the test are carefully developed to be as clear and unambiguous as possible. For example, lower-frequency words and less common syntactic structures aren't included unless they're part of the construct tested.

- **Diversity.** As this test is taken by students of diverse backgrounds, the question pool includes substantial content that visibly reflects U.S.-based racial and ethnic diversity as well as international or global perspectives, cultures, and settings. Questions also reflect a balanced representation of genders.

Content Specifications for the ACCUPLACER ESL Language Use Test

Overall Claim for the Test

The ESL Language Use Test is intended to collect evidence in support of the following claim about student performance:

> *ELLs can demonstrate their developed ability to identify and apply learned conventions of standard written English grammar, usage, and punctuation to a range of English sentence structures on a variety of topics.*

Test Description

The ESL Language Use Test measures students' developed ability to use correct grammar in English sentences. Specific grammatical conventions measured are nouns/pronouns/pronoun case structure; sentence structure (including prepositions, parallelism, and sentence fragments/run-ons); subject-verb agreement; adverbs/adjectives (including modifiers and comparatives); verbs (e.g., verb tenses, regular/irregular verb forms); and subordination/coordination.

Questions on the Language Use Test come in two formats. Half the questions require test takers to make a sentence grammatically correct by filling in a blank with a word or phrase from the choices provided. The remaining questions present test takers with two sentences and ask them to select from the choices presented the sentence that best combines them while maintaining the meaning and intent of the two discrete sentences.

Visit the ACCUPLACER practice website to see sample questions: **https://accuplacer.collegeboard.org/student/practice**.

Table 3.13 provides a synopsis of key content dimensions of the test.

Table 3.13: *ACCUPLACER ESL Language Use Test Content Specifications*

Content Areas	Number of Questions	Percentage of Test*
Question types		
Fill in the black	10–10	50-50
Sentence combining	10–10	50-50
Content Assessed		
Nouns, pronouns, pronoun case structure	2–3	10–15
Sentence structure	4–5	20–25
Subject-verb agreement	2–2	10–10
Adjectives/Adverbs	2–3	10–15
Verbs	4–4	20–20
Subordination/Coordination	3–4	15–20
Total	20	100

* Percentages don't necessarily add up to 100 because some questions meet multiple content requirements.

Key Features of the ESL Language Use Test

Distinctive features of the ESL Language Use Test are described below:

- **Skills that matter most.** The ESL Language Use Test focuses on test takers' developed ability to apply critical standard English conventions in short, one-sentence contexts in the service of meaning and correct form. Test takers must also demonstrate the developed ability to paraphrase, combining sentences as appropriate in order to improve delivery and flow, reduce redundancy, and avoid repetitive words, phrases, and structures.

- **Accessible contexts.** As a test for students representing a wide range of cultural and linguistic backgrounds, experiences, and lengths of exposure to U.S. culture and English-speaking environments, the ESL Language Use Test uses contexts carefully developed to be as accessible and free of elements that would impede access to the content as possible. Prior topic-specific knowledge, including knowledge of U.S. culture or norms, isn't tested.

- **Accessible language.** While language isn't simplified when it's part of the construct being assessed, passages and questions used on the test are carefully developed to be as clear and unambiguous as possible. For example, lower-frequency words and less common syntactic structures aren't included unless they're part of the construct tested.

- **Diversity.** As this test is taken by students of diverse backgrounds, the question pool includes substantial content that visibly reflects U.S.-based racial and ethnic diversity as well as international or global perspectives, cultures, and settings. Items also reflect a balanced representation of genders.

Content Specifications for the ACCUPLACER ESL Listening Test

Overall Claim for the Test

The ESL Listening Test is intended to collect evidence in support of the following claim about student performance:

> *ELLs can demonstrate their developed ability to understand ideas and information presented in connected discourse on a variety of topics, compensating for gaps in their lexical and structural control of the language by using contextual clues.*

Test Description

The ESL Listening Test measures the listening skills of nonnative speakers of English. Test takers listen to a series of connected discourses, including lectures and conversations involving multiple speakers, and answer questions that assess

their developed ability to comprehend the information and ideas conveyed. Listening segments that relate to academic environments include lectures and conversations that take place at various locations on a school or campus (e.g., a computer lab or library) and may reference academic content students are expected to encounter in college. They may include exchanges among students or between a student and a professor or college administrator. Such exchanges typically reflect language that college students can expect to experience either on campus or in social interactions, such as common idioms and appropriate tone and register. Other conversations may reflect those that occur in everyday environments, such as interactions that may happen at a restaurant, a doctor's office, at work, or at social events. Less challenging listening segments typically contain familiar contexts, structures, and contexts, high-frequency words and phrases, and clear underlying organization. More challenging ones tend to be longer; focus on more complex discourses on less familiar topics and situations using less familiar vocabulary; and tend to use a wider and less predictable range of grammatical, syntactic, and organizational structures.

On the test, test takers are presented with pictures of the speakers and then listen to a conversation or lecture. They then hear questions related to the lecture or conversation and pick the best answer from four options. Questions assess two listening skills. Literal Comprehension questions measure developed abilities to identify the main subject of a conversation or lecture, the ability to identify details and explicitly stated information and the ability to understand instructions. Implied Meaning questions include those that assess the comprehension of relationships between ideas such as cause-effect and comparison-contrast; the developed ability to identify speakers' purpose, tone, and point of view; and the developed ability to make reasonable inferences and predictions based on the information heard.

Visit the ACCUPLACER practice website to see sample questions: **https://accuplacer.collegeboard.org/student/practice**.

Test Summary
Table 3.14 provides a synopsis of key content dimensions of the test.

Table 3.14: *ACCUPLACER ESL Listening Content Specifications*

Content Areas	Number of Questions	Percentage of Test*
Question types	8–12	40–60
Literal comprehension	8–12	40–60
Implied meaning		
Passage types		
Conversation	14–16	70–80
Lecture	4–6	20–30
Total	20	100

Key Features of the ESL Listening Test
Distinctive features of the ESL Listening Test are described below.

- **Skills that matter most.** The ESL Listening Test focuses on core listening skills ELLs need for access to and success in college-level courses, including skills that aid understanding of information that is expressed explicitly and that which is implied in utterances and spoken discourse on a range of topics and in both academic and social contexts.

- **Accessible contexts.** As a test for students representing a wide range of cultural and linguistic backgrounds, experiences, and lengths of exposure to U.S. culture and English-speaking environments, the ESL Listening Test uses contexts carefully

developed to be as accessible and free of elements that would impede access to the content as possible. Prior topic-specific knowledge, including knowledge of U.S. culture or norms, isn't tested.

- **Accessible language.** While language isn't simplified when it's part of the construct being assessed, passages and questions used on the test are carefully developed to be as clear and unambiguous as possible. For example, lower-frequency words and less common syntactic structures aren't included unless they're part of the construct tested.

- **Diversity.** As this test is taken by students of diverse backgrounds, the question pool includes substantial content that visibly reflects U.S.-based racial and ethnic diversity as well as international or global perspectives, cultures, and settings. Items also reflect a balanced representation of genders.

Direct measures

The ACCUPLACER essay tests, WritePlacer and WritePlacer ESL, provide a direct measure of test takers' writing skills. Scores can be used in conjunction with other ACCUPLACER test scores to assist with the placement of students into college level, developmental or corequisite, and/or English as a second language courses.

WritePlacer

WritePlacer is intended to collect evidence in support of the following claim about student performance:

Overall Claim for the Test

> *Students can demonstrate their developed ability to successfully create an original essay in which they use reasoning, personal experience, observations, readings, and an appropriate rhetorical approach to effectively develop a point of view or position on an issue.*

Test Description

WritePlacer is designed to help college administrators assess students' readiness to successfully meet postsecondary writing demands. Test takers are directed to write an original essay in which they develop a point of view on an issue using reasoning and evidence to support their ideas.

The writing task is presented through a prompt. WritePlacer prompts are designed to enable students to react and respond quickly in a variety of ways. Prompts are easily accessible to a range of student populations, including ESL students. Each prompt comprises a short, single-paragraph passage and a writing assignment based on the passage. Test takers are encouraged to respond to the prompt based on their personal experiences, observations, studies, or readings; to accept or reject the ideas presented to whatever extent they see fit; and to use the rhetorical approach that best helps them fulfill the writing task. The test doesn't reward or punish formulaic approaches to writing, such as the five-paragraph essay. Test takers aren't judged on the positions they take, and the essay is viewed as a draft and evaluated as such.

Visit the ACCUPLACER practice website to see a sample prompt: **https://accuplacer.collegeboard.org/student/practice**.

Test Summary

Table 3.15 provides a synopsis of key dimensions tested on the WritePlacer.

Table 3.15: *WritePlacer Dimensional Score Descriptions*

Dimensions and Descriptions
Purpose and focus
The extent to which the writer presents information in a unified and coherent manner, clearly addressing the issue. Specific elements considered include unity, consistency, coherence, relevance, and audience.
Organization and structure
The extent to which the writer orders and connects ideas. Specific elements considered include introduction, thesis, paragraphs, transitions, and conclusions.
Development and support
The extent to which the writer develops and supports ideas. Specific elements considered include point of view, coherent arguments, evidence, and elaboration.
Sentence variety and style
The extent to which the writer crafts sentences and paragraphs demonstrating control of vocabulary, voice, and structure. Specific elements considered include sentence length, sentence structure, usage, tone, vocabulary, and voice.
Mechanical conventions
The extent to which the writer expresses ideas using Standard English Conventions. Specific elements considered include spelling, grammar, and punctuation.
Critical thinking
The extent to which the writer communicates a point of view and demonstrates reasoned relationships among ideas. Specific elements considered include clarity, depth, precision, logic, accuracy, breadth, and relevance.

Key Features of the WritePlacer

Distinctive features of the WritePlacer are described below:

- **Skills that matter most.** Students taking WritePlacer are scored on their command of writing skills that would be required in most college courses, including skills for planning and presenting information and ideas in a coherent piece of writing, use of precise language, logical presentation of ideas, development of a point of view, and clarity of expression.

- **Accessible contexts.** As an essay test for students representing a wide range of cultural and linguistic backgrounds, experience, and length of exposure to U.S. culture and English-speaking environments, WritePlacer uses contexts carefully developed to be as accessible and free of elements that would impede access to the content as possible. WritePlacer prompts are made up of short passages, don't test specialized, technical, or literary topics, and test takers aren't asked to respond in a way that requires knowledge of U.S. culture or norms.

- **Accessible language.** Care is taken to make sure WritePlacer prompts test writing achievement and that the reading required to understand and fulfill the task is as clear and unambiguous as possible.

WritePlacer ESL

WritePlacer ESL is intended to collect evidence in support of the following claim about student performance:

Overall Claim for the Test

> *ELLs can demonstrate their developed ability to successfully create an original essay in which they draw on a broad range of experiences, learning, and ideas to effectively develop a point a view or position on a given topic.*

Test Description

WritePlacer ESL provides a direct measure of the writing skills of students who identify as ELLs. Test scores help colleges determine the appropriate level of English instruction to provide either before or while students take college-level courses.

On WritePlacer ESL, test takers write an essay in response to a prompt on a familiar idea or topic. Prompts are short, clearly and simply stated, and designed to allow test takers of diverse backgrounds and writing abilities to tap into their own experiences, interests, and perspectives in order to produce viable writing samples for evaluation.

Each WritePlacer ESL prompt is made up of a short text of no more than three sentences and a writing assignment that focuses the writing task.

Visit the ACCUPLACER practice website to see a sample prompt:
https://accuplacer.collegeboard.org/student/practice

Test Summary

Table 3.16 provides a synopsis of key dimensions assessed on the WritePlacer:

Table 3.16: *WritePlacer ESL Dimensional Score Descriptions*

Dimensions and Descriptions

Word Use

The extent to which the writer is able to use a wide range of words and phrases accurately. Specific elements considered include range of vocabulary, appropriateness and accuracy of word choice, use of idiomatic expressions, choice of word forms (parts of speech) appropriate to context, and spelling.

Sentence Use

The extent to which the writer is able to use a variety of sentence patterns with both independent and dependent clauses. Specific elements considered include sentence boundaries (avoidance of fragments, run-ons, comma splices), control of noun, verb, and adjective clause structure, sentence variety, word order, punctuation and capitalization.

Grammar

The extent to which the writer is able to express ideas using grammatically correct English. Specific elements considered include verb forms, verb tenses, pronoun usage, agreement, count/noncount nouns, gerunds/infinitives, prepositions, and articles.

Organization and Development

The extent to which the writer is able to focus on the assigned topic and to develop ideas clearly. Specific elements considered include focus, organization, and support.

Key Features of WritePlacer ESL

Distinctive features of the WritePlacer are described below.

- **Skills that matter most.** Students taking WritePlacer ESL are scored on the writing skills college-bound ELLs must develop in order to adequately prepare to participate in college-level courses using English as the medium of instruction and evaluation. These include skills for fulfilling the writing task in a coherent and focused way using appropriate words and phrases, a variety of sentence structures, and conventional grammar.

- **Accessible contexts.** As a test for students representing a wide range of cultural and linguistic backgrounds, experience, and length of exposure to U.S. culture and English-speaking environments, WritePlacer ESL prompts use contexts that are carefully selected and are as accessible and free of elements that would impede access to the content as possible. Prompts are short passages, don't test specialized, technical, or literary topics, and test takers aren't asked to respond in a way that requires knowledge of U.S. culture or norms.

- **Accessible language.** Care is taken to make sure WritePlacer prompts test writing achievement and that the reading required to understand and fulfill the task is as clear and unambiguous as possible.

3.3. Development of the ACCUPLACER Suite of Assessments
Item Specifications

Except for WritePlacer and WritePlacer ESL, which elicit student writing using prompts, all questions in the ACCUPLACER Suite are multiple-choice, with each question having one and only one correct or best answer. Some items are discrete in format, while others are set-based. Prompts are accessible in content and comparable in the challenge posed so that they provide a consistent and reliable measure of test taker achievement.

In keeping with Standard 4.7 (AERA, APA, & NCME, 2014), the following section describes how College Board creates and reviews the multiple-choice questions and prompts in the ACCUPLACER Suite of Assessments.

Crafting of Questions and Tasks

Test and Question/Task Specifications

Given the defined test domains, College Board measurement and content staff worked with educational experts to prepare test and question/task specifications that represent the depth and breadth of the defined domains and help ensure the consistent development of assessments of the highest quality. The specifications define the question/task types and formats required to measure most directly and authentically the domains of skills and knowledge relevant to the ACCUPLACER Suite of Assessments' primary purposes and the tests' overall claims.

Passage Selection and Question/Task Design

The ACCUPLACER Suite measures durable skills and knowledge needed in postsecondary education, work, and life. All content area tests are developed to elicit from students work worth doing through questions and tasks aligned with the best classroom practices. College Board does this by working with a multitude of teachers in K–12 and postsecondary instructors of entry-level courses across the United States.

Reading Test

In the ACCUPLACER Reading Test, students engage with texts worth reading and worthy of careful consideration. Some passages are selected from previously published authentic writing that exemplifies the genres represented on the test and are designed to be powerful, insightful pieces; others are commissioned passages of high quality. The essential first step of question development is a close and careful reading of the text. Test questions resemble questions that might emerge naturally in a thoughtful classroom conversation and return students to the text to examine closely the information and ideas within it. The best test questions develop out of a sensitive engagement with the passage rather than an effort to try to cover in a mechanical way every possible testing point in the domain. They also favor a more organic development process that respects the unique natures of rich texts in a variety of content areas.

Writing Test

The ACCUPLACER Writing Test is composed of passages that are engaging and challenging, paired with questions that focus clearly on a core of writing and language skills and knowledge empirically linked to college and career readiness and success

requirements. These commissioned passages are designed to provide meaningful contexts for the skills and knowledge being addressed and exemplify the qualities of effective arguments, informative/explanatory texts, and nonfiction narratives. Test questions assess writing and language skills and knowledge in extended prose contexts rather than in isolation and require students to make active choices in revision and editing rather than simply identify errors.

Math Tests

The three ACCUPLACER Math Tests (Arithmetic, QAS, and AAF) ask students to demonstrate their command of the math most provably useful in a range of college courses and career environments. They provide the opportunity for richer applications of the most essential math to address real-world situations and problems. In QAS and AAF in particular, these core topics are examined extensively and at a high level of proficiency.

Test questions are thoughtfully designed with the help of educators with a deep knowledge of the target mathematical content and practices. The questions on each Math Test emphasize the use of math in unlocking insights and solving problems. The test design allows the core of math to be examined with the range of rigor required (as defined through evidence) to be college and career ready, examining at once students' procedural skill, application, and conceptual understanding. Rather than covering a broad number of topics that most students will never see again, students study fewer topics that represent a deep core that they can draw upon again and again in their schooling, college, and career. At the same time, the assessment includes pure math problems that focus on the type of reasoning essential for success in solving diverse problems and engaging in demanding disciplines.

ESL Tests

In the ACCUPLACER multiple-choice ESL tests (Reading Skills, Sentence Meaning, Language Use, and Listening), students engage with material worth reading, worthy of consideration, reflective of the range of cultural experiences of ELLs, and appropriate for ELLs of varying degrees of achievement and exposure to the target language and culture. Passages included on the Reading Skills and Listening Tests are commissioned and typically represent content that the ELL population is likely to be able to access following a close and careful consideration of the content presented. Test questions return students to the material to examine closely the information and ideas within it and typically focus on core reading, writing, grammar, and listening skills and knowledge, so that assessment results provide meaningful information on what English skills test takers must develop in order to feel confident using English so that independent learning may continue, to succeed in college-level courses, to acquire content delivered in English, and to communicate successfully in an English-speaking educational environment.

WritePlacer and WritePlacer ESL Prompt Development

WritePlacer and WritePlacer ESL prompts are written by test developers who are subject-matter specialists, many of whom have either high school or college teaching experience. They are then reviewed by writing faculty members representing high schools and two-year and four-year colleges from around the United States. Each prompt is written to be easily accessible to the test-taking population, including students for whom English is a second language and students from a range of age groups.

Crafting of Prompts

WritePlacer

As WritePlacer is intended to give test takers the opportunity to use a broad range of experiences, learning, and ideas to support their point of view on the issue addressed, prompts don't draw on specialized knowledge in any particular area or on any specific course material that a student may have studied. They're free of figurative, technical, or

specific literary language or references and are highly varied and drawn from a range of fields and reflective of a range of student interests, including the arts, sports, technology, science, and history.

In light of the above, each prompt presents a passage that engages a broad range of students, stimulates critical reflection, and allows test takers to draw on their knowledge and interests to respond. Each prompt comprises a short passage of no more than 80 words, followed by a writing assignment that focuses the student on the issues addressed in the passage and presents the writing task. Passages used are based on previously published texts and are selected based on their utility for the task and their appropriateness and suitability for a wide audience. As much as possible, passages are kept intact as they originally appeared in publication, although they may be adapted to provide greater accessibility for students (e.g., to ensure fairness or eliminate unduly obscure or difficult vocabulary or construction). Each passage is followed by a prompt-specific assignment that succinctly states the issues presented in the prompt and identifies possible points to consider as students plan and execute their essay.

All prompts are written to meet a set of criteria: they must present an issue that will engage test takers from a broad range of backgrounds and allow them to draw on their knowledge and interests to respond; they must stimulate critical reflection on the issue by suggesting a range of possible viewpoints within it (i.e., passages may present opposing points of view on an issue, but each side should be complex enough to allow the student to develop a variety of positions within each point of view); they must avoid moralizing statements that might invite expected or hypocritical responses; they mustn't be flat statements of fact; and they mustn't ask students to address a particular audience (e.g., their high school principal) or write a certain kind of response (e.g., a letter to the mayor), as these might make test takers think there is one socially or politically appropriate way to address the prompt, thus potentially impeding the writing activity these prompts are developed to elicit.

In keeping with AERA/APA/NCME Standard 3.2, test developers work to ensure that scores aren't influenced by construct-irrelevant characteristics (AERA, APA, & NCME, 2014). Prompts are reviewed by College Board as well as independent content matter experts familiar with the testing population to ensure that they're fair and don't disadvantage any particular subgroup in the student population based on factors such as race, ethnicity, gender, sexual orientation, or socioeconomic status.

Sample WritePlacer prompts are available at **https://accuplacer.collegeboard.org/student/practice**.

WritePlacer ESL
WritePlacer ESL, like other ESL tests in the ACCUPLACER Suite, is intended to support the placement of college-bound ELLs into appropriate language instruction courses. For WritePlacer ESL, an ESL essay test, to meet this purpose, each prompt used to elicit writing samples from test takers must be written to be accessible to ELLs with a range of writing skills and degree of exposure to the U.S. language and culture. Specifically, while a prompt needs to contain enough context to introduce the task and focus the writing activity, it's intentionally short and uncomplicated so that gaps in reading skills won't disqualify test takers from responding or prevent them from demonstrating their developed abilities to think about, develop a perspective about, and write about the topic. Each prompt is free of figurative or technical language or specific literary references. Instead, each is written to give test takers the opportunity to draw on a wide range of experiences, learning, and viewpoints to produce a writing sample that supports the test taker's point of view on a topic. The prompts require test takers to assume a specified role or address a specific audience or task that may encourage the perception of, or put them in a position of, having to exhibit familiarity with appropriate social or political norms or with acceptable behaviors in contexts they may not be able to access.

Each prompt comprises a short statement of one to three sentences and a writing assignment that focuses the test taker on the statement and the writing task. All prompts are written to satisfy a set of criteria: they must incorporate all references needed to respond; be of interest to students and support the generation of a 300- to 600-word essay; be clearly stated and unambiguous, comply with conventions of grammar and usage and spelling; contain no language that reinforces stereotypes; and be free of content that could be judged offensive to members of any group. The goal of the WritePlacer ESL prompt is to elicit a writing sample that demonstrates the range of developed abilities ELL test takers have acquired, thus enabling the identification of their need for further writing instruction, either in an ESL writing course or independently.

In keeping with AERA/APA/NCME Standard 3.2, test developers work to ensure that scores aren't influenced by construct-irrelevant characteristics (AERA, APA, & NCME, 2014). Prompts are reviewed by College Board as well as independent content matter experts familiar with the testing population to ensure that they're fair and don't disadvantage any particular subgroup in the student population based on factors such as race, ethnicity, gender, sexual orientation, or socioeconomic status.

Sample WritePlacer ESL prompts are available at
https://accuplacer.collegeboard.org/student/practice.

Question/Task Writer Support Materials
In order to consistently develop assessments with engaging, rich stimulus materials and contexts that lend themselves to high-quality questions and tasks, College Board has developed and continues to maintain a range of test support materials intended to help make sure that all questions and tasks are evidence-based, valid, and accessible to all students—in short, that they meet the highest possible standards (AERA, APA, & NCME, 2014). These materials include question/task writer content and fairness guidelines as well as question/task prototypes and templates. College Board contracts with faculty and educational professionals at both the high school and postsecondary levels and with other independent content and instructional experts to develop and/or review all questions and tasks. In this way, those most familiar with the student population of interest and knowledgeable in the best instructional practices in the field make the most significant contribution to assessment content. This helps ensure that the test materials and tasks included in the assessments are engaging, instructionally appropriate, and fair to all students.

Question Content and Fairness Reviews
Multiple-Choice Questions
Prior to pretesting, all stimuli, questions and tasks are reviewed by external, independent reviewers who are asked to evaluate each question and task according to a set of criteria for content soundness and fairness. In keeping with Standard 4.8, these reviewers are typically active classroom teachers drawn from across the nation from both the secondary and postsecondary levels and are deeply familiar with the student population of interest and the nature and purpose of the test (AERA, APA, & NCME, 2014).

Content reviewers are focused on ensuring the soundness of each question, task, and stimulus and evaluating its relationship to the construct (e.g., reading) being measured, its relevance and appropriateness to the work students do in high school, and its value in terms of measuring students' degree of college and career readiness. Fairness reviewers are charged with helping ensure that test questions, tasks, and stimuli are broadly accessible to the wide-ranging student population that takes the assessment, that the questions and tasks are clearly stated and unambiguous in their intent, and that the questions and tasks don't offer unfair advantages or disadvantages to some students. See Section 2.3 for more information on the fairness review.

Prompts

As noted previously, texts are written or selected for use in prompts based on their appropriateness and suitability for a wide audience. Once passages have met internal criteria for selection based on topic, length, and text complexity, secondary and postsecondary classroom teachers then review the prompts for any issues of fairness. In keeping with AERA/APA/NCME Standard 3.2, the test developers work to ensure that scores aren't influenced by construct-irrelevant characteristics (AERA, APA, & NCME, 2014). Prompts are reviewed to ensure that they don't advantage or disadvantage any particular subgroup in the student population based on factors such as race/ethnicity, gender, sexual orientation, or socioeconomic status. College Board is careful to select topics that test takers are likely to be able to respond to by drawing on their experiences but are, at the same time, unlikely to cause emotional responses that may get in the way of the writing task. For more information on Fairness Reviews prior to testing, see Section 2.3.

Each WritePlacer and WritePlacer ESL prompt is reviewed by test developers as well as by a writing test development committee to maximize clarity and to eliminate ambiguity. It's further reviewed for fairness by an independent panel of educators that reflects the diversity of WritePlacer ESL test takers. External panels are representative groups made up of high school and college faculty from a variety of academic disciplines, geographic regions, and races/ethnicities. Panels are also balanced for gender. During the review and discussion process, prompts may be discarded, accepted, or revised to eliminate ambiguity, improve the wording, etc.

Item Pretesting, Analysis, and Calibration

Item Pretesting

Multiple-Choice Questions

Every operational question in the ACCUPLACER Suite of Assessments has previously been pretested; that is, the question has been embedded in an operational test and administered to students (not for a score) in the target population to make sure that it isn't ambiguous or confusing and to determine the difficulty level and the degree to which it differentiates among higher- and lower-achieving test takers. The pretest responses are also analyzed to determine whether test takers of different racial/ethnic groups or genders, having similar achievement levels, respond to the question differently.

For each pretested question, the data from at least 1,000 test takers are used to evaluate question performance. This information provides an accurate estimate of how the question will function when administered operationally.

Prompts

In keeping with AERA/APA/NCME Standard 4.8, it's important to ensure that the prompts "function similarly for different groups" (AERA, APA, & NCME, 2014, p. 88). After prompt reviews and a resolution process to address any concerns raised during the reviews and review discussions finish, new essay prompts are field tested[14] with a representative sample of test takers in a special administration in classrooms around the country. For each group of prompts field tested, a diverse sample of schools is invited to participate by having students respond to a particular prompt. The students who participate in field testing vary by race/ethnicity, gender, and socioeconomic status. A sample of at least 300 responses to each essay prompt is obtained to enable the analysis of the performance of the new prompts and to determine whether they're accessible to students and to provide exemplars of various levels of writing competence for use in the scoring process described in Chapter 5.

---

14. ACCUPLACER questions may be pretested or field tested. Multiple-choice ACCUPLACER questions are pretested; they are embedded in operational forms and administered to students but are not included in their scores. WritePlacer and WritePlacer prompts are field tested; they are administered outside of operational tests to students who resemble the intended populations.

The responses gathered from field tests are read by a group of experienced writing instructors to determine whether a particular prompt is readily understood by test takers and elicits responses that reflect differing degrees of writing skill. In other words, does the prompt lead to responses that can be scored validly and that exhibit differentiation among able and less able writers? Members of this group individually read and score a substantial number of the responses. As a group, they discuss each prompt and decide whether it's usable, needs revision, or should be discarded. From the student responses collected during the field testing, exemplars are chosen for each point on the holistic scoring scale. These serve as anchor papers for training essay readers[15] when the essay prompt is administered operationally. The scoring process is described in Chapter 5.

Analysis of Pretest Information

In keeping with Standard 4.10, data collected from pretests are analyzed to provide important information about the appropriateness of items to be included in the item pools of the ACCUPLACER Suite of Assessments (AERA, APA, & NCME, 2014). An initial item analysis is performed on the data to provide test developers with statistical information to review items for any possible issue with regard to keys and distractors, as well as an alert for possible issues that will affect item calibrations. The main statistics computed are indices of difficulty, discrimination, and differential item functioning (DIF).

Item Difficulty and Discrimination. Initial analysis of difficulty and discrimination is based on Classical Test theory (CTT). In CTT, item difficulty is the percentage of test takers who answer the item correctly. It is typically referred to as the "p-value" of the item. A high p-value indicates an easy item; that is, an item that most of the test takers answered correctly. A low p-value indicates a hard item, one that most of the test takers answered incorrectly. Item discrimination is the correlation between the score of an item and the total scores of a test. Items that correlate well with total test score tend to correlate well with one another and produce a test that is more reliable. A high correlation indicates that the item performs as expected, in that the proportion of higher-scoring test takers answering the item correctly is greater than the proportion of lower-scoring test takers answering correctly. A low correlation indicates an item not performing as intended and requires a review by test development experts and may need to be removed from the test.

When an item is dichotomously scored, point-biserial correlation is equivalent to a well-known Pearson correlation coefficient to indicate the item-total correlation coefficient to indicate item-total correlation. The computation of point-biserial index is shown in the equation below. There, $\bar{x}_{i1}$ is the mean scale score for test takers who answered item i correctly, $\bar{x}_{i0}$ is the mean scale score for test takers who answered item i incorrectly, p_i is the proportion of test takers that answered item i correctly, and s_x is the standard deviation of scale scores:

$$r_{pbs} = \frac{(\bar{x}_{i1} - \bar{x}_{i0})\sqrt{p_i(1 - p_i)}}{s_x}$$

In addition to the difficulty and discrimination indices above, other statistics computed at this stage are:

- The number and percentage of test takers who selected each distractor
- The point-biserial correlation for each distractor
- The average scale score for test takers who selected each distractor

---

15. WritePlacer and WritePlacer ESL essays are electronically scored by the Intelligent Essay Assessor (IEA). Those essays that cannot be scored by the IED (off-topic, too short, written in a language other than English) are read and scored by human readers. Readers also routinely audit machine-scored essays.

CTT item statistics are used to flag items for a closer examination. The following are the criteria used to flag items for further content review:

- Item Difficulty < 0.15; Item Difficulty > 0.90
- Item Discrimination: Item Discrimination < 0.10

Additionally, items were also flagged based on the performance of the following distractors:

- Distractor Discrimination: > 0.05
- Distractor Attracting < 1% of all test takers
- Distractor Average Scale Score Higher than that for the Keyed Response Option

Items with incorrect key, unclear distractors, extreme p-values, or low item-total correlations were dropped from calibration.

Differential Item Functioning (DIF)

Establishing the fairness of tests is an important part of supporting and justifying the use of test scores for their intended purposes. Of particular concern in establishing test fairness is ensuring that items are equally informative for different subgroups of test takers. For test scores to be valid, it is important to be certain that there is nothing influencing responses to items other than the knowledge, skills, and abilities (KSAs) that the items on the test intend to measure (Zumbo, 1999). Anything unrelated to the intended KSAs that differentially influences the responses of subgroups of test takers is a threat to the validity of score interpretations. In short, when subpopulations of test takers are matched on their abilities, there should be no difference in their achieving a particular score on test items. To identify any potentially unfair test items, a statistical technique known as Differential Item Functioning (DIF) analysis can be employed.

DIF is a statistical observation that involves matching test takers from different groups on the characteristic measured and comparing performance across groups on each item. Test takers of equal ability who belong to different groups should respond similarly to a given test item. If they do not, the item is said to function differently across groups and is classified as a DIF item (see Clauser & Mazor, 1998, or Holland & Wainer, 1993 for more complete descriptions of DIF theory and methodology). Differential performance does not mean an item is biased. Item bias is present when an item has been statistically flagged for DIF and the reason for the DIF is traced to a factor irrelevant to the construct the test is intended to measure. Therefore, for item bias to exist, a characteristic of the item that is unfair to one or more groups must be identified. ACCUPLACER test items flagged for DIF are sent to test developers for review.

For analysis of DIF for gender, the performance of male test takers is compared to the performance of female test takers, with the males serving as the reference group. For analysis of DIF for ethnic/racial groups, the performance of White test takers as the reference group is compared to other ethnic/racial subgroups. Ethnicity is defined as Hispanic or non-Hispanic, and race is defined as American Indian or Alaska Native (AIAN), Asian, Black or African American, Multiple Races, and White. All non-Hispanic respondents are identified as one of the previously listed racial categories. The minimum sample size requirements are 50 for the focal group and 100 for the reference group when calculating the statistics.

There are many methods for detecting DIF (Clauser & Mazor, 1998; Camilli & Shepard, 1993; Holland & Wainer, 1993). Most of them are not applicable for computer-adaptive tests (CATs), since CAT doesn't have the same set of items, or even the same length of the test, for all test takers. Logistic regression (LR) is one of the methods that is appropriate for CAT DIF detection for the following advantages (Sireci, 2001; Swaminathan & Rogers, 1990; Zumbo, 1999). First, LR doesn't require test takers to take the same set of items. Second, LR can detect both uniform and non-uniform DIF. Uniform DIF occurs when the probability of getting an item correct is higher for one group across the ability level. Nonuniform DIF occurs when the probability of getting an item

correct is higher at one range of ability level but lower at the other range for one group. Third, simulation studies have shown that LR has acceptable power and type I error rate when employing the effect size measure (Jodoin & Gierl, 2001). DIF analyses for the ACCUPLACER Suite of Assessments employ Zumbo's 1999 method.

In this method, three models are constructed for each item. Accordingly, the R^2, which is based on the likelihood ratio c^2 for testing the null hypothesis that all coefficients are 0 (Cox & Snell, 1989), can be obtained for each. The effect sizes for uniform and non-uniform DIF are $R_2^2 - R_1^2$ and $R_3^2 - R_2^2$, respectively. The models and effect sizes are in in Table 3.17. For the ACCUPLACER Suite of Assessments, an effect size is considered:

- Negligible if it is less than 0.034
- Moderate if it is greater than or equal to 0.034 and less than 0.07
- Sizeable if it is greater than or equal to 0.07

Items with sizeable effect sizes, those items favoring one group over the other for test takers of the same ability, are not included in the calibration. Such items may be revised and re-pretested. Items with moderate effect sizes are reviewed to determine whether some aspect of what the item is measuring is particularly related to subgroup membership and irrelevant to the construct being measured. When an item is identified as having such characteristics, they are also excluded from calibration. They may be revised or re-pretested.

Table 3.17: *Models and Effect Sizes*

Index	Model	R^2	Effect Size
1	$y = b_0 + b_1 theta$	R_1^2	
2	$y = b_0 + b_1 theta + b_2 group$	R_2^2	$R_2^2 - R_1^2$
3	$y = b_0 + b_1 theta + b_2 group + b_3 theta * group$	R_3^2	$R_3^2 - R_2^2$

Item Pool Calibration

Items that are not dropped from the item pool based on the initial item analysis and subsequent review are included in the item calibration. Item calibration is the term commonly used to describe Item Response Theory (IRT) item parameter estimation. IRT is fully described in the CAT Algorithm section of this chapter and it is suggested that, as the details of IRT and parameter estimation are beyond the scope of this manual, interested readers refer to Hambleton and Swaminathan (1985), Hambleton, Swaminathan, and Rogers (1991), Lord (1980), Lord & Novick (1968), and Baker and Kim (2004). Excellent discussions of IRT within the context of computer-adaptive testing can be found in Wainer (2000).

IRT calibrations for the Arithmetic, QAS, AAF, Reading, and Writing items were performed using FlexMIRT® Version 3.51 (Cai, 2017). Items for the three Math Tests were calibrated concurrently. All ESL items were calibrated using BILOG-MG version 3.0 from Scientific Software, Inc. (Zimowsky, Muraki, Mislevy & Bock, 1996). After the first calibration, item model-fit was inspected. In some cases, items were eliminated prior to final calibration. The item parameters and the item response functions were examined for abnormalities. For an item to be selected, it must demonstrate the ability to distinguish between high proficiency and low proficiency test takers. Items not rejected after final calibration become part of the respective item pools.

3.4. Computer-Adaptive Test Algorithm

Nine of the 11 tests in the ACCUPLACER Suite of Assessments are administered as a computer-adaptive test (CAT), which allows for instantaneous score reporting. The technology in CAT affords the capability to provide accurate and efficient measurement

of a test taker's knowledge and skills. As soon as a test taker finishes a test, his or her Individual Score Report is available and is immediately exportable into existing campus information systems.

The previous sections in this chapter discussed the creation and testing of items to be administered in ACCUPLACER placement tests. As the last few sections discussed the establishment of an item pool, this section will explain the workings of the CAT Algorithm for the ACCUPLACER Suite of Assessments. In keeping with AERA/APA/NCME Standard 5.16, "when test scores are based on model-based psychometric procedures, such as those used in computerized adaptive or multistage testing, documentation should be provided to indicate that the scores have comparable meaning over alternate sets of items" (AERA, APA, & NCME, 2014, p. 106).

As in any CAT, the adaptive algorithm used for the ACCUPLACER tests is designed to arrive at reliable scores as efficiently as possible. The rationale for this approach is that it is unnecessary for test takers of high ability to take the easiest items or for test takers of low ability to take the hardest items, as doing so does not contribute much to the quality of the ability estimates. Items with difficulty levels that are far away from a test taker's ability level do not contribute enough information about estimation of the test taker's ability to be of practical use. Choosing items that do contribute information is more efficient and provides more accurate scores.

In a typical CAT, a test taker is initially presented with an item of a designated difficulty level. In some testing programs, the first item is of medium difficulty, while in ACCUPLACER the first item is of a slightly easy difficulty to allow a positive introduction to the testing experience. If the test taker's response to the first item is correct, a more difficult item is presented next. If the response is incorrect, the test taker is presented with a less difficult item. An ability estimate based on the test taker's previous responses is computed after each response and successive items are presented to meet content specifications and to provide as much information as possible about the test taker's ability. The test is terminated after the specific number of items are administered.

A typical CAT system consists of several components:

- An item response theory (IRT) model
- A calibrated item pool
- An initial trait level to begin the test
- A procedure for selecting items
- A method for estimating ability
- A criterion to terminate the test

The rest of this section is organized according to these components.

Item Response Theory (IRT) Model

Constructs such as ability in a subject matter are not directly observable. The responses that test takers provide to items related to the construct to be assessed provide information about the unobserved construct. The test development process, from defining the construct of interest to generation, review, and field testing of the items, ensures that the items presented operationally to a test taker are appropriate representations.

Tests are sets of items that are an operationalization of the construct of interest. IRT models are intended to relate the estimated response to the underlying construct. IRT models describe a probabilistic relationship between a test taker's response on a test item and some latent trait such as math, reading, or writing ability. Test takers with higher ability have a higher probability of answering an item correctly than test takers with lower ability. Figure 3.1 depicts this relationship; as test takers increase in ability, as indicated on the X-axis, the probability of answering the item correctly increases, as indicated on the Y-axis. The figure is an Item Characteristic Curve (ICC) for a specific item based on the three-parameter logistic model that is used for the ACCUPLACER placement tests.

As all of the items in the ACCUPLACER Suite of Assessments non-essay tests are multiple-choice, a dichotomous model is most appropriate. The most general of the common dichotomous models is the three-parameter logistic model. The three parameters are discrimination (known as a-parameter), difficulty (known as b-parameter) and pseudo-guessing (known as c-parameter). Referring to Figure 3.1, the a-parameter is proportional to the slope of the ICC at the difficulty level of the item (b, discussed below). The steeper the slope, the more discriminating the item is performing around that difficulty level. That is, with a higher slope it does not require a large change in ability to increase the probability of answering correctly. The probabilities of answering an item correctly increase more slowly for a less-discriminating item. Typically, it is desirable to have items with a-values of 1 or higher, but content constraints and the difficulty of creating items with high discriminations at differing ability levels generally means that items with a-values lower than 1 are often used.

The pseudo-guessing parameter, c, represents the probability of answering the item correctly for test takers with extremely low or no knowledge of the construct. For multiple-choice tests, it is possible to answer an item correctly by guessing.

The difficulty parameter, b, occurs at the ability level where the test takers with that level have a certain probability of answering correctly. If there were no guessing, the b-parameter would be at the ability level where there was a 50% probability of answering the item correctly. When there is possibility of guessing, the b-parameter would be at the ability where the probability of answering the item correctly is equal to one-half of the sum of 1 and the guessing parameter c; $(1 + c_i)/2$.

Theta (q) is the ability level on the underlying and unobservable trait being measured. The range of theta is theoretically from negative infinity (absolutely no knowledge) to positive infinity (perfect knowledge). Though the scale for ability and difficulty parameters is arbitrary, most IRT software scales the test taker parameters so that 0 is the average, and that the standard deviation of the abilities is generally set to 1. This means that an item with a b-parameter of 0 is usually considered to be of average difficulty. For most purposes, the range of θ is between -5 and +5 and for other purposes ranges from -3 to +3.

The three-parameter model is represented by the following equation:

$$P_i(\theta) = c_i + (1 - c_i)\frac{e^{1.7a_i(\theta - b_i)}}{1 + e^{1.7a_i(\theta - b_i)}}$$

where Pi(θ) is the probability of a correct response to item i , given an ability level of θ. The item parameters are ai, bi, and ci and refer to characteristics of the items themselves.

Figure 3.1. Graphical Representation of Item Characteristic Curve
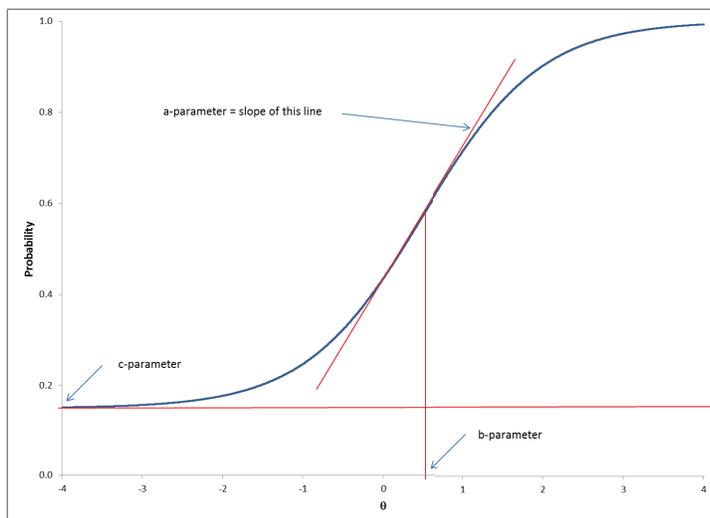
Other popular IRT models are special cases of the more general three-parameter logistic model. The two-parameter model is the case where the c-parameter is set equal to zero. The one-parameter model is obtained when the c-parameter is zero and the a-parameter is set equal to 1 for all items.

The calibrations performed with IRT have the result that all test takers' theta values and all item parameters are on scale. That is, by virtue of all the items being on the same scale each test form (i.e., selection of items administered to a student) is effectively pre-equated. Thus, post-equating (which is what most people call equating) is unnecessary.

Calibrated Item Pool

With requirements and constraints that need to be satisfied for a successful administration of a CAT, a rich item pool is essential. For each ACCUPLACER CAT, a pool of test items written to the various test content areas are developed and, following extensive review including pretesting, are calibrated to the selected IRT models. The number of items on each test question pool is presented in Tables 3.18–3.26 below.[16] The distribution of items across content areas covered in each test is also presented. Other statistical characteristics of the test question pools are in Appendix E: Statistical Characteristics of ACCUPLACER Test Question Pools. These include distribution information of item parameters, and item parameter classification by content constraints.

Table 3.18: *ACCUPLACER Reading Test Question Pool Content Distribution*

Content Areas	Number of Questions	Percentage of Pool
Information and Ideas	109	40.22
Rhetoric	104	38.38
Synthesis	20	7.38
Vocabulary	38	14.02
Total	271	100.00

Table 3.19: *ACCUPLACER Writing Test Question Pool Content Distribution*

Content Areas	Number of Questions	Percentage of Test*
Expression of Ideas	197	59.70
▪ Development	65	19.70
▪ Organization	65	19.70
▪ Effective Language Use	67	20.30
Standard English Conventions	133	40.30
▪ Sentence Structure	29	8.79
▪ Conventions of Usage	37	11.21
▪ Conventions of Punctuation	67	20.30
Total	330	100.00

Table 3.20: *ACCUPLACER Arithmetic Test Question Pool Content Distribution*

Content Areas	Number of Questions	Percentage of Pool
Whole number operations	40	6.71
Fraction operations	78	13.09
Decimal operations	172	28.86

(Continued)

16. Tables 3.18–3.26 represent the content distributions as of February 25, 2019.

Content Areas	Number of Questions	Percentage of Pool
Percent	165	27.68
Number comparisons and equivalents	141	23.66
Total	596	100.00

Table 3.21: *ACCUPLACER Quantitative Reasoning, Algebra, and Statistics Test Question Pool Content Distribution*

Content Areas	Number of Questions	Percentage of Pool
Rational numbers	140	8.47
Ratio and proportional relationships	164	9.92
Exponents	141	8.53
Algebraic expressions	224	13.55
Linear equations	441	26.68
Linear applications and graphs	139	8.41
Probability and sets	63	3.81
Descriptive statistics	90	5.45
Geometry concepts for Pre-Algebra	127	7.68
Geometry concepts for Algebra 1	124	7.50
Total	1,653	100.00

Table 3.22: *ACCUPLACER Advanced Algebra and Functions Test Question Pool Content Distribution*

Content Areas	Number of Questions	Percentage of Pool
Linear equations	441	22.45
Linear applications and graphs	139	7.08
Factoring	176	8.96
Quadratics	191	9.73
Functions	165	8.40
Radical and rational equations	299	15.23
Polynomial equations	169	8.61
Exponential and logarithmic equations	59	3.00
Geometry concepts for Algebra 1	124	6.31
Geometry concepts for Algebra 2	101	5.14
Trigonometry	100	5.09
Total	1,964	100.00

Table 3.23: *ACCUPLACER ESL Reading Skills Test Pool Content Distribution*

Content Areas	Number of Questions	Percentage of Pool
Skills assessed		
Literal Comprehension	91	49.19
Inference	94	50.81
Passages support either one or two questions	147	79.46
One-question passage	38	20.54
Two-question passage		
Content areas		
Science	46	24.86
History/Social Science	23	12.43
Arts/Humanities	39	21.08
Narratives	33	17.84
Psychology/Human Relations	23	12.43
Practical Situations	21	11.35
Total	185	100.00

Table 3.24: *ACCUPLACER ESL Sentence Meaning Test Pool Content Distribution*

Content Areas	Number of Questions	Percentage of Pool
Question types		
Fill in the blank	142	53.99
4-option multiple-choice	121	46.01
Content Assessed		
Particles, phrasal verbs, prepositions	48	18.25
Adverbs, adjectives, connectives sequence	85	32.32
Basic nouns and verbs	82	31.18
Common idioms	48	18.25
Total	263	100.00

Table 3.25: *ACCUPLACER ESL Language Use Test Pool Content Distribution*

Content Areas	Number of Questions	Percentage of Pool
Question types		
Fill in the black	168	62.45
Sentence combining	101	37.55
Content Assessed		
Nouns, pronouns, pronoun case structure	41	15.24
Sentence structure	59	21.93
Subject-verb agreement	28	10.41
Adjectives/Adverbs	32	11.90
Verbs	55	20.45
Subordination/Coordination	54	20.07
Total	269	100.00

Table 3.26: *ACCUPLACER ESL Listening Test Pool Content Distribution*

Content Areas	Number of Questions	Percentage of Pool
Question types		
Literal comprehension	276	51.98
Implied meaning	255	48.02
Passage types		
Conversation	450	84.75
Lecture	81	15.25
Total	531	100.00

Initial Trait Level

For each placement test, the first question is chosen based on a relatively low initial ability estimate of $\theta = -1.0$ to allow most test takers a successful experience in the beginning of the test. From the second question on, the next question administered to a test taker is automatically chosen based on the skill level indicated by answers to all previous questions and the content specifications that are still to be met.

Item Selection Procedure

ACCUPLACER tailors the test to each test taker using an item-selection algorithm that takes into account content balance, measurement precision, and item exposure control. Content balancing is an important consideration that ensures that different tests across test takers cover the same proportion of content categories so that test takers are measured on the same composite of traits. The adaptive nature of the tests involves identifying and administering the items in the pool that provide the most information at the current estimate of ability for each test taker as they progress through the test. Item exposure control is another important practical consideration. Because CATs are continuously administered from the same item pool over a period of time, without control for item exposure, some "popular" items may become known and no longer provide valid measurement. To ensure appropriate content coverage, efficient and accurate scores, and to prevent overexposure of items, the ACCUPLACER CAT system incorporates statistical algorithms to control content balancing, information, and item exposure rate for the item selection process.

The exposure control algorithm used for ACCUPLACER tests is based on the Conditional Randomesque method (Kingsbury & Zara, 1989). This method allows a preset maximum exposure rate for a specific ability range so that exposure will be constrained at various ability levels. For all placement tests, the preset maximum exposure rate is 0.25.

The Conditional Randomesque strategy randomly selects the next item to be administered from the group of the most informative items, given the current theta estimate. The selection of the item is always made at random among the most informative items. The Conditional Randomesque repeatedly selects the same number of the most informative items (e.g., 2, 3, 4…10) from which one is randomly selected for administration throughout testing and does not switch to maximum information selection at any time. Kingsbury and Zara suggest that continuing the randomization technique throughout testing will decrease the overlap in items seen by test takers of similar abilities. The number of most informative items from which one will be selected for administration to the test taker is 4 for all ACCUPLACER tests. These group sizes were found to work well for present maximum exposure rate of 0.25.

To ensure content balance, a number of constraints are built into each test with respect to content category and inter-item dependency. Inter-item dependencies deal with relations of exclusion and inclusion between items in the pool (Veldkamp & van der Linden, 2000). As part of inter-item dependency, constraints can be specified for "enemy"

items such that the presentation of one item will preclude the appearance of another item in the enemy list on the same test, or for "set" items such that all items in the same set will be administered together on the test.

To balance content, the ACCUPLACER CAT algorithm uses the adjusted Weighted Penalty Function Method (Segall and Davey, 1995; Fan, 2007) which takes into account the content-related constraints, item information, and sufficiency of items related to each constraint. This method assigns each eligible item in the pool a penalty value at each item selection level, with items having smaller penalty values deemed more desirable for selection. After the penalty value for each item is computed, a list of items with the lowest penalty values is formed and provided to the item exposure control method for further item selection. This ensures that the next item to be administered is selected according to content constraints, maximum test information, and the desired item exposure rate. The content categories for each test and the number of items selected from each category are presented in the test content tables in this chapter.

Ability Estimation Procedure

Although fewer questions are presented with a CAT for each test than would be given in a linear test, greater accuracy is maintained on trait estimates by providing challenging tests that correspond to each test taker's skill level, and by using an ability estimation algorithm that ensures accurate and efficient ability estimates. For ability estimation, the Maximum Likelihood Estimation procedure that employs the Newton-Raphson method and Brute Force method (Birnbaum, 1968; Hambleton & Swaminathan, 1985) is implemented in the ACCUPLACER CAT algorithm. The Newton-Raphson method is the most commonly used method in IRT pattern scoring; it has significant advantage in efficiency but has two misuses that can occur for some test takers. Some sequences of item responses from a test taker may be inconsistent with expectations (e.g., answering some harder items correctly and easier items incorrectly) resulting in an inability of the method to converge on an estimate of ability. For other patterns, the method may find a local maximum estimate when, in fact, another real maximum exists elsewhere on the ability scale. The Brute Force method is an algorithm often used in Computer Science for searching sets of data for an answer. It requires more extensive computation but guarantees a true maximum likelihood solution. For the ACCUPLACER tests, the Newton-Raphson method is used as the main ability estimation method, with the Brute Force method as a supplement when the convergence and local maximum issues occur. Estimated abilities are in the -5.0 to 5.0 range of scores. The reported scores for each are based on a linear transformation from the ability estimates. How the linear transformation is derived is discussed in the Scaling section of Chapter 6: Psychometrics.

Termination Criterion

ACCUPLACER placement tests are all *fixed-length* tests consisting of dichotomously scored multiple-choice items. The number of items on each test is presented in the Test Specifications section of this chapter. In a *variable-length* CAT, which is used by some testing programs, the test terminates when the precision of the ability estimate for a test taker reaches an established threshold. Because ACCUPLACER tests have a predetermined number of items to be administered for each test, a variable-length CAT termination rule was not used.

Other Features of the CAT Algorithm

The ACCUPLACER platform also allows items to be pretested by embedding the pretested items, which do not count toward test takers' scores, into the computer-adaptive test. In addition to the operational computer-adaptive test algorithm, there is a computer-adaptive test simulation system built for ACCUPLACER that has the exact same computer-adaptive test functionalities as the operational system. The simulation system is capable of simulating computer-adaptive tests and analyzing the characteristics of the simulated tests to provide information about the performance of the computer-adaptive test algorithm. For the current computer-adaptive test algorithm, extensive simulation studies have been conducted; these studies indicate that the

current system produces tests that meet content and item exposure requirements sufficiently well and provide ability estimates that are psychometrically efficient and accurate, resulting in valid and reliable test scores.

3.5. Accommodations

ACCUPLACER Accommodations

College Board believes that, in keeping with AERA/APA/NCME Standards, "all test takers should have the full opportunity to demonstrate their standing on the construct being measured" (AERA, APA, & NCME, 2014, p. 52). Consistent with the Americans with Disabilities Act and to ensure fairness across assessments, students who present documentation that their disabilities affect their ability to participate in the ACCUPLACER Suite of Assessments are eligible for accommodations. Approval from College Board isn't needed to administer accommodations on ACCUPLACER assessments; institutions work with their SSD coordinators to determine eligibility based on test takers' documented disabilities.

COMPANION Tests

College Board is committed to making assessments available in accessible formats. All tests in the ACCUPLACER Suite, with the exception of the ESL Listening Test, have two corresponding, comparable COMPANION forms. COMPANION tests present ACCUPLACER content in alternate formats and are designed for test takers who aren't able to take computer-adaptive tests or for institutions that may be unable to administer them. These tests are available in several formats: digital, nonadaptive versions that may be taken on the computer; print-on-demand versions that test administrators may download from the platform; audio CDs; braille; and regular and large print booklets. COMPANION tests are linear tests that have been designed to proportionally align in content to the corresponding computer-adaptive tests. They use the same score scale as the computer-adaptive tests and may be up to two times the length of their corresponding computer-adaptive tests. Table 3.27 shows the lengths of COMPANION tests relative to the computer-adaptive tests.

Table 3.27: *Number of Questions for Computer-Adaptive and COMPANION Tests*

Test	Number of Questions	
	Computer-Adaptive	COMPANION
Reading	20	40
Writing	25	40
Arithmetic	20	40
Quantitative Reasoning, Algebra, and Statistics	20	40
Advanced Algebra and Functions	20	40
ESL Reading Skills	20	35
ESL Sentence Meaning	20	35
ESL Language Use	20	35
ESL Listening	20	n/a
WritePlacer	1	1
WritePlacer ESL	1	1

In keeping with AERA/APA/NCME Standards, COMPANION tests yield inferences that are comparable to those from the computer-adaptive version of the assessment (AERA, APA, & NCME, 2014).

Development of COMPANION Tests

The ACCUPLACER COMPANION Tests are developed to provide an alternate, paper-based test to test takers or institutions that may be unable to access the computer-based test. AERA/APA/NCME Standard 8.3 states that "when the test taker is offered a choice of test format, information about the characteristics of each format should be provided" (AERA, APA, & NCME, 2014, p. 134).

The COMPANION Tests are designed to have content specifications that align with the ACCUPLACER computer-adaptive tests and to provide good measurement across the score scale. Two parallel COMPANION forms are available to test takers for every test. The two versions of each COMPANION form have similar reliabilities, which is to say that both forms are built to have comparable Test Information Functions (TIF). Figures E-1–E-8 in Appendix E: Statistical Characteristics of ACCUPLACER Test Question Pools show the TIF and Test Characteristic Curves (TCC) of each COMPANION form.

The COMPANION forms are created using only items selected from active item pools. The College Board Psychometrics group is responsible for selecting the initial draft of items used in each test form to meet content and statistical specifications. The form is sent to the College Board's Assessment Design and Development (AD&D) group for review. The AD&D review may result in a decision to replace an item or items based on their professional judgment. AD&D sends their feedback to Psychometrics, who make any changes, including item replacement, if necessary and check that statistical specifications are still met. The form is then sent back to AD&D for additional review to assure that the test meets content specifications. This process is repeated until both test forms meet psychometric and content specifications. During this process, AD&D will also determine the order that the items will be presented on the final test form.

Development of Conversion Tables

The raw scores that are computed for the COMPANION tests are the sum of the correct answers, the "number right." These raw scores are then converted to the reported score scale, which is 200 to 300 for the Reading, Writing, Arithmetic, QAS, and AAF, and 20 to 120 for the four ESL CATs.

The conversion tables for the COMPANION forms were developed in order for the "number correct" scores from each COMPANION test to be placed on the appropriate score scale. The analyses involved simulation studies using Item Response Theory (IRT) item parameters for the items of each COMPANION form. The IRT parameters are described further in the section on CAT Algorithm. A sample of 50,000 simulees with a uniform distribution of ability (theta) ranging from very low ability (theta equal to -5.00) to very high ability (theta equal to 5.00) is used for the simulation in order to ensure uniform coverage across the ability range.

First, these sampled thetas are used as true theta values as input to the simulations, and the simulation steps are as follows:

1. For each COMPANION form, student response data, strings of incorrect (0) and correct (1) responses, were simulated for each simulee via the three-parameter item response theory (IRT) model, using the theta values and item parameters for the items on the COMPANION test.

2. The total "number correct" score is then computed for each simulee on each form of each test by summing the zeroes and ones simulated in step 1. The "number right" ranges from zero to the total number of items on the test. Note: Scores as low as zero are generally not observed in simulations or in operational administrations.

3. The estimated theta for each response string is then computed using IRT.

4. At each number correct score from zero to number of items on the test, the estimated theta values for all simulees achieving that score were summarized and the average estimated theta is obtained. Theta is estimated using expected a posteriori (Bock & Mislevy, 1982).

5. The average true theta is then mapped to the scale score using the theta-to-scale score conversion table established for the computer-adaptive test of each subject area.

As a result, the ACCUPLACER scale score is obtained for the complete range of number correct scores on each COMPANION form. When a COMPANION test is administered, the number correct is computed and the reported score is generated by finding the scaled score associated with that number in the conversion table.

Testing Requirements

College Board works to ensure that all test scores are valid for their intended uses and that all test takers have a fair testing experience. This chapter documents the appropriate use of all ACCUPLACER assessments, how they should be administered, and the steps all administrators of these tests must take to protect test materials and prohibit the inappropriate sharing of test information at any time.

Access to ACCUPLACER is granted to nationally or internationally accredited, degree-granting institutions and system educational governance. Applications by nonprofit organizations will be reviewed on a case-by-case basis. Institutions or organizations that don't meet the established criteria for becoming an ACCUPLACER user may appeal the decision to the ACCUPLACER Program. The final decision regarding access lies with the ACCUPLACER Program.

Section 4.1 of this chapter discusses the appropriate use of ACCUPLACER as an assessment to help place students in higher education courses, as well as the process taken to develop ACCUPLACER test items that are fair and eliminate bias. Section 4.2 discusses the policies and procedures involved in ACCUPLACER test administration, the environment in which these tests should be administered, and also provides an overview of the eligibility and responsibilities of those administering the tests. Section 4.3 discusses test security and the ways to prevent attempts to gain an unfair advantage and compromise test scores for their intended uses. We discuss these procedures as they apply to test materials and test takers, as well as the rationale behind these procedures.

4.1 Appropriate Use

ACCUPLACER English and Math placement tests are designed to be administered to entering undergraduate students. Their primary function is to determine students' college and career readiness and inform college placement. ACCUPLACER English as a Second Language (ESL) tests are intended for use in placing nonnative speakers of English into ESL courses.

Appropriate use of ACCUPLACER in high schools

While designed to inform college placement decisions, ACCUPLACER may be administered to identify high school students who are eligible for dual enrollment. The administration of ACCUPLACER to high school students is also appropriate to the extent that scores from the tests are used to evaluate the college and career readiness of these students. In these instances, the high school using it as such would be one working in partnership with higher education institutions or departments of education, and it would have an established college and career readiness definition and metric. ACCUPLACER tests would be administered to these high school students late in the junior year or early in the senior year so they have time to work toward meeting the college's requirements.

Inappropriate use of ACCUPLACER in high schools

ACCUPLACER should not be used as a high school graduation requirement, accountability assessment, or as a tool for evaluating teaching and learning.

4.2 Test Administration

Administration of Online Tests

All ACCUPLACER tests, except for WritePlacer and WritePlacer ESL, are computer-adaptive. Computer-adaptive testing is a mode of test administration that uses computer algorithms to select and deliver test items to test takers. In an adaptive test, a pool of test items has been written for the various content areas of each test.

Each ACCUPLACER test question, or item, has been calibrated for difficulty and other characteristics. Unlike many traditional tests where all test takers take a single form of an assessment, the sequence of test questions and the questions themselves will vary from test taker to test taker. The next question administered to a test taker is automatically chosen to yield the most information about the test taker based on the skill level indicated by answers to all prior questions. The criteria for selecting the next item to be administered to a test taker are complex; however, the primary criterion is a desire to match the difficulty of the item to the test taker's current estimated proficiency.

The ACCUPLACER test delivery system adapts or "tailors" the test to each test taker by keeping track of a test taker's performance on each test item and using an item-selection algorithm based on a weighted deviations model to determine the next item to be administered. During testing, the first question presented is of medium difficulty and is chosen randomly from several starter questions of the same level of difficulty. If a test taker answers the question incorrectly, the next question to be administered is chosen from a group of easier questions. If the test taker answers the question correctly, the next question presented will be somewhat more difficult.

The test delivery system continues this process throughout the test, choosing the next question that is expected to yield the most information about the test taker. To assure that the test is balanced in content, and that the kinds of questions presented don't differ greatly from one test taker to another except in difficulty, several constraints are built into the program. These constraints guide the selection of questions to be administered so that a balance is achieved regardless of the skill level of the individual.

Administration of Accommodated Tests

COMPANION tests provide accommodated formats for test takers who are unable to take computer-adaptive ACCUPLACER tests. COMPANION tests are available for all ACCUPLACER tests except for the ESL Listening Test. In addition to the regular paper-and-pencil format, COMPANION tests are also available in braille and large print, and digital. Audio CDs that test takers can use to hear ACCUPLACER test stimuli, questions, and answer choices are also available.

The COMPANION tests typically have 1.5 to 2 times as many items as the corresponding computer-adaptive tests. Items for the COMPANION Tests are selected using an Automated Test Assembly program. The program follows the same content specification for the adaptive tests, resulting in test forms that are proportionally equivalent in content coverage. The program also uses statistical specifications to ensure that selected items cover the difficulty range from easy to difficult. Additional information on COMPANION test administration can be found in the *COMPANION Administrator's Manual* (**https://accuplacer.collegeboard.org/pdf/accuplacer-companion-manual.pdf**)*.*

4.3 Security

Institutions using ACCUPLACER are required to sign a License Agreement that requires all testing be done in a secure and proctored setting. The agreement further requires that test takers be monitored at all times during a test session; that an authorized, certified test administrator from the institution be present on-site during all administrations of ACCUPLACER or COMPANION tests; that only approved users may log in to the ACCUPLACER platform to administer an assessment; that test takers not be permitted to log in to the platform on their own; and that under no circumstances should proctor login credentials be shared with test takers. Login credentials may not be written on chalkboards or whiteboards, printed, emailed, or presented online in any form or place. The ACCUPLACER User's Guide (**https://accuplacer.collegeboard.org/pdf/accuplacer-user-guide.pdf**) contains more information about login options.

Test Center Guidelines

Before administering the ACCUPLACER tests, administrators should evaluate their testing facilities and review testing procedures so as to ensure a comfortable, positive, and efficient testing environment and experience for test takers. Recommended testing

facilities include college or university test centers or testing offices, or a quiet, proctored room in high schools. Examples of prohibited sites include a test taker's home[17], place of employment, coffee shops, and internet cafes.

Below are mandatory guidelines for any ACCUPLACER testing environment:

1. The testing room must be appropriately heated or cooled, adequately ventilated, and free from distractions.

2. Lighting must enable all test takers to read the computer screen in comfort and shouldn't produce shadows or glare on the computer screen or writing surfaces.

3. The testing room cannot contain maps, periodic tables, posters, charts, or any ancillary materials related to the subject matter of the test.

4. The testing room must comfortably accommodate the number of testing stations placed in it.

5. Computer stations must be set up to restrict visibility of other computer screens in the testing room.

6. Each test station should include a computer monitor, keyboard, and mouse that are properly positioned for ease of use, as well as a comfortable chair with a back.

7. Testing rooms must be quiet throughout the duration of each test administration. When testing is scheduled or is in progress, other activities that would disrupt the standardized testing environment cannot be conducted.

8. The building, testing rooms, and restrooms should be accessible to people with disabilities, including wheelchair accessibility.

9. Restrooms should be located near the testing room and should be easy to find. Post directional signs if necessary.

10. Unauthorized individuals (e.g., parents, chaperones, non-testing test takers) are not permitted in the testing center during the test. Persons authorized to assist for accommodation purposes (e.g., reader or scribe), are permitted to be present during testing.

Prohibited Items

1. Any non-medical electronic devices, especially any device capable of recording audio, photographic, or video content or any device capable of viewing or playing back such content. This includes but is not limited to cellular phones, tablets, pagers, smartphones, walkie-talkies, PDAs, wireless communication devices; digital cameras, digital watches, smartwatches, or wristwatch cameras; listening devices such as radios, media players (with or without headphones), or recorders; and flash/thumb drives or any other portable electronic storage or recording device.

2. Any unauthorized testing aids, including calculators (unless for a prescribed accommodation); test taker provided keyboard, computer, or laptop, unless there is a documented disabling condition that requires the use of such specific device; dictionaries (standard and/or bilingual), books, pamphlets, or other reference materials; and slide rules, protractors, compasses, or rulers.

3. Paper of any kind (scratch paper may be provided, and any scratch paper distributed for testing must be collected and destroyed by the test center administrator; test takers cannot use or discard their own scratch paper).

4. Food, beverages, or tobacco products.

5. Weapons, firearms, or other items prohibited by law or test center/campus safety and security policies.

---

17. Test takers with or without documented disabilities are permitted to test at home provided they use one of ACCUPLACER's approved virtual proctors. Before providing off-campus proctors (i.e., proctors not affiliated with ACCUPLACER's virtual testing network) with proctor credentials, the referring institution must thoroughly vet the requested proctor's identity and employment to ensure that assessments are being administered in accordance with ACCUPLACER requirements. Failure to do so can result in suspension of your ACCUPLACER account.

Proctor eligibility and responsibilities

Proctors must meet the following eligibility criteria:

1. Proctors must review the ACCUPLACER proctor training materials and pass the ACCUPLACER Certificate of Test Administration (ACTA).

2. Proctors must be responsible adults trained to administer standardized tests.

3. High school students are ineligible to proctor ACCUPLACER.

4. Proctors must have their own username and password. Login credentials cannot be shared with anyone, including Institution Administrators (IAs) and Site Managers (SMs).

5. Proctors cannot administer an ACCUPLACER test to a member of their household, immediate and/or extended family members, or friends.

6. Proctors must not have a stake in the outcome of test takers' scores.

7. Proctors cannot be engaged with any commercial test preparation company. This includes employment, volunteering, consulting, or acting as independent contractors.

8. For any remote, off-campus location testing, proctors must be vetted and authorized by the institution to proctor assessments in such locations.

Proctors are eligible to receive proctor login credentials only after they have successfully passed the ACTA. These credentials are valid for one year and need to be renewed on an annual basis. IAs and SMs must select proctors who are trained in the administration of standardized tests, which includes how to safely secure all ACCUPLACER testing materials (online and COMPANION). In addition, IAs and SMs are expected to provide proctors with specific information about test administration procedures, as well as regular training. All parties involved in administering ACCUPLACER tests must adhere to the policies outlined in the ACCUPLACER License Agreement (found in the ACCUPLACER Program Manual at **https://secure-media.collegeboard.org/digitalServices/pdf/ accuplacer/accuplacer-program-manual.pdf**).

Proctors must engage in active proctoring behavior. For example, they should circulate the testing room throughout the testing session to ensure that test takers are working on the correct test and not engaging in any prohibited behavior. During the administration of a test, proctors cannot engage in non-test administration activities, such as reading, eating, drinking, conversing, or using cell phones or other electronic devices.

Proctor responsibilities vary and include the following:

1. Verify the identity of every test taker before the administration of a test. In the event a test taker leaves the testing center for any reason during testing, identification must be re-verified upon their return to the testing center or upon receiving their Individual Student Report (ISR).

2. Collect test takers unauthorized items (e.g., cellphones, smartwatches, dictionaries, etc.) and during the test store them in a secure area that is accessible to the test taker. Test takers cannot place these within arm's reach (e.g., under their desks or chairs).

3. Support the IA and/or SM with securing all ACCUPLACER tests.

4. Assist test takers with testing equipment during testing and/or with logging in to the correct ACCUPLACER test.

5. Provide test takers with scratch paper and pencils and collect and securely destroy all scratch paper once testing is completed. Test Takers cannot bring or use their own scratch paper.

6. Print and distribute ISRs to test takers after testing. Identification must be re-verified prior to providing an ISR to a test taker at the end of the test session.

7. Administer assessments to test takers with disabilities if they are familiar with their specific accommodation needs.

8. Ensure proper test security in advance of, during, and following testing sessions.

Interpretation and Application of Result

To ensure that scores are usable for intended purposes, "assessment instruments should have established procedures for test administration, scoring, and interpretation." (AERA, APA, & NCME, 2014, p. 114). Test administration procedures were discussed in Chapter 4. The first section of this chapter (Section 5.1) describes the scoring procedures for the ACCUPLACER placement tests and their COMPANION forms as well as WritePlacer. The chapter then covers scale score reporting (Section 5.2). Because ACCUPLACER scores are used to make placement decisions, procedures for setting placement scores are discussed in Section 5.3. To help students and institutions in interpreting performance on the tests, ACCUPLACER Skills Insight, which is discussed in Section 5.4, provides a set of data-driven statements of what students know and can do at different ranges of performance on the tests. Because high stakes decisions such as college placements should not be based on a single test score, a discussion of the use of multiple factors concludes the chapter (Section 5.5).

5.1 Scoring Procedures
ACCUPLACER Placement Tests
Generating test scores for a test taker involves several steps in a computer-adaptive test (CAT) and is covered in Chapter 3. When a test taker provides an answer to a question, a key is applied to ascertain whether the answer was correct or not. The test taker's ability estimate is computed, and an appropriate item chosen to be administered. This sequence is followed until the last item is administered. The discussion of the process of choosing appropriate items and calculation of ability estimate is in the section on CAT algorithm, also in Chapter 3. The ability estimate is then translated to the reported score scale, as discussed in Chapter 6.1: Scaling Procedures. The reported score is then ready for reporting to the test taker and institutions.

COMPANION Forms
The raw scores that are computed for the COMPANION tests are the sum of the correct answers, or the "number right." These raw scores are then converted to the reported score scale which is 200 to 300 for the Reading; Writing; Arithmetic; Quantitative Reasoning, Algebra, and Statistics (QAS); and Advanced Algebra and Functions (AAF) tests, and 20 to 120 for the ESL tests using the conversion tables.

The conversion tables for the COMPANION forms were developed so the "number correct" scores from each COMPANION test could be placed on the appropriate score scale. The analyses involved simulation studies using Item Response Theory (IRT) item parameters for the items of each COMPANION form. The IRT parameters are described in the CAT Algorithm section of Chapter 3. A sample of 50,000 simulees with a uniform distribution of ability (theta) ranging from very low ability (theta equal to -5.00) to very high ability (theta equal to 5.00) is used for the simulation to ensure uniform coverage across the ability range.

First, these sampled thetas are used as true theta values as input to the simulations. The simulation steps are as follows:

1. For each COMPANION form, student response data, the strings of incorrect (0) and correct (1) responses, were simulated for each simulee via the three-parameter item response theory (IRT) model, using the theta values and item parameters for the items on the COMPANION test.

2. The total "number correct" score is then computed for each simulee on each form of each test by summing the zeroes and ones simulated in Step 1. The "number right" ranges from zero to the total number of items on the test. Note: Scores as low as zero are generally not observed in simulations or in operational administrations.

3. The estimated theta for each response string is then computed using IRT.

4. At each number correct score from zero to the number of items on the test, the estimated theta values for all simulees achieving that score are summarized and the average estimated theta is obtained. Theta is estimated using the expected a posteriori (EAP) method (Bock & Mislevy, 1982).

5. The average true theta is then mapped to the scale score using the theta-to-scale score conversion table established for the computer-adaptive test of each subject area.

As a result, the scale score is obtained for the complete range of number correct scores on each COMPANION form. When a COMPANION test is administered, the number correct is computed and the reported score is generated by finding the scale score associated with that number in the conversion table.

WritePlacer and WritePlacer ESL Essay Scoring

The WritePlacer and WritePlacer ESL tests measure students' ability to write effectively, which is critical to academic success. Each test taker's writing sample is scored on the basis of how effectively it communicates a whole message to readers for the purpose stated in the question. The test taker's score is based on their ability to express, organize, and support opinions and ideas, not the position taken on the essay topic.

Each WritePlacer or WritePlacer ESL essay gives test takers an opportunity to show how effectively they can develop and express their ideas in writing. They read a short passage and an assignment question that are focused on an important issue and then write an essay in which they develop their own point of view on the issue. The passage is intended to stimulate thought about a topic or issue and test takers are asked to draw on a broad range of experiences, learning, and ideas to support their point of view on the issue in question.

WritePlacer tests are machine scored. Human participation only comes in at three steps:

- Scoring of essays used for initial calibration of the scoring engine
- Annual WritePlacer audit
- Checking of student essays that have anomalies and need to be referred to human readers

College Board provides the WritePlacer scoring guide, which is an eight-point holistic scale incorporating criteria that characterize writing. The scoring guide for WritePlacer ESL is a six-point scale. The writing sample is evaluated holistically. Scores on WritePlacer range from 1 to 8, and scores for WritePlacer ESL range from 1 to 6. An essay that is too short to be evaluated, written on a topic other than the one presented, or written in a language other than English will be given a score of zero. The WritePlacer and WritePlacer ESL scoring rubrics can be found in Appendix A: Scoring Rubrics-WritePlacer and WritePlacer ESL.

In addition to holistic scores, additional feedback for each student essay is available in the form of dimension scores on:

- Purpose and focus
- Organization and structure
- Development and support
- Sentence variety and style
- Mechanical conventions
- Critical thinking

WritePlacer ESL essays receive dimension scores on:

- Word use
- Sentence use
- Grammar
- Organization and development

Dimensional scoring rubrics are also available in Appendix A. The score on each dimension has three levels. The dimensional scores are not meant to sum up to the holistic score for the essay. They are only meant to provide additional feedback on the test taker's writing ability.

Technology Used to Score WritePlacer Essays

WritePlacer essays are electronically scored by the Intelligent Essay Assessor (IEA) that is powered by the Knowledge Analysis Technologies (KAT) engine. Developed by the Knowledge Technologies group at Pearson, the IEA is a unique automated assessment technology that evaluates the meaning of text, not just grammatical correctness or spelling.

The IEA is based on Latent Semantic Analysis (LSA), a statistical language learning theory and computer model that measures the semantic similarity of words and documents with accuracy closely approximating that of human judges. LSA was originated at Bell Laboratories under Thomas Landauer, Ph.D., and was built into automated educational assessment products at the University of Colorado and Pearson.

The IEA automatically evaluates the semantic substance of a student's writing by comparing a new essay to a set of essays that have been graded by two expert human readers. It can do this comparison and produce accurate and reliable scoring because each essay question has been calibrated against 500 or more student essays with human scores.

As a new essay is submitted, the IEA looks for similarities to the scored essays and assigns a holistic score by placing it in a category with the essays to which it is most similar. Dimension scoring occurs in much the same way. For each dimension, the system assesses the student essay by comparing it to scored essays, and then categorizes the dimension in question. The IEA includes built-in detectors for off-topic responses and other special situations that may need to be referred to human readers.

The correlation and agreement rates of the scores produced by the IEA have been shown to be as high as or higher than those between two independent readers.

Calibrating the Scoring Engine

Rater Training

Human raters score the essays that are used to calibrate the IEA. As noted above, each essay question is calibrated against 500 or more student essays scored by human raters. Raters are people with at least a bachelor's degree who have qualified to score the assessments that have been brought in to assign scores. The current scoring pool is experienced in scoring overall (eight or more years) and has scored WritePlacer for at least four years.

All rater training for WritePlacer is conducted online. WritePlacer scorers attend an online training program that teaches the fundamentals of holistic and analytic scoring. The training is designed so that scorers learn how to properly apply the rubrics. Scorers must pass a baseline qualification test in order to score WritePlacer essays. Scorers also receive training for every WritePlacer prompt that they score. There are two baseline trainings, one for WritePlacer prompts, and one for WritePlacer ESL prompts. Scorers initially train and qualify on the WritePlacer prompt, which qualifies them to score all WritePlacer prompts.

Baseline training consists of 1 anchor set, 2 practice sets, and 2 qualifying sets. There are also 3 online training modules that every scorer takes. Outlier scorers also take a 4th module unique to WritePlacer that is used to train how to score the types of outlier responses that cannot be scored by the KAT scoring engine, as well as how to monitor workflow and manage their own item assignments to meet the 24-hour scoring deadline.

A key component of the training is the anchor papers, which contain exemplar student responses that clarify the scoring guide and define the range that exists within each score point. The anchor papers demonstrate different approaches and different levels of achievement within each score point. Anchor sets are accompanied by annotations, which explain the score of each anchor paper. Examples selected from the student response help explain the score. Ultimately, the annotations help the scorer understand not only that particular paper, but also similar papers seen during training and scoring.

Scorers are required to get at least 50% exact and 90% exact plus adjacent agreement on 1 of 2 qualifying sets on a baseline item. The qualifying standard is applied independently to each scoring trait. After qualifying, prompt specific training is available for every standard WritePlacer prompt. Prompt specific training is 1 anchor set followed by a single practice set.

Some scorers also train and qualify on the baseline item for WritePlacer ESL. ESL baseline training uses three of the same modules as WritePlacer training, so scorers do not need to retake the modules for ESL scoring. However, they do need to take a Scoring WritePlacer ESL module. There are also 1 anchor set, 2 practice sets, and 2 qualifying sets in this training. Each ESL prompt also has prompt specific training that consists of 1 anchor set and 1 practice set.

ACCUPLACER uses a 60-day requalification rule. If a scorer goes more than 60 days without scoring at least one response, they must requalify.

Process for Obtaining Scored Anchor Papers for Training and Scoring Engine

Anchor papers from administration to administration are reused since this project uses the same prompts year to year. Anchor sets for all prompts were created using live responses from outlier scoring. Sets were built by content experts with over 10 years of experience working on writing assessments. Anchor papers are generated from live outlier essays.

Anchor sets cover the 8-point holistic score range as well as the trait scores, but the primary training focus is on the holistic score. All anchor papers are annotated.

Human Rater Validation of Scores for a Sample of Essays

As a quality measure to ensure consistency of scores between human scorers and the rubric, a variety of techniques to monitor scoring quality are administered.

Backreading is a primary tool for proactively guarding against scorer drift. The scoring system's integrated backreading tool allows their supervisory staff to review the scores assigned to individual student responses by any given scorer.

Scoring directors and supervisors can perform a search for:

- Responses scored by a particular scorer
- Responses receiving a particular score point
- Responses with scores that agree with, are adjacent to, or are nonadjacent to each other
- Combinations of these features

Scoring directors use calibration sets to reinforce rangefinding standards, introduce scoring decisions, or correct scoring issues and trends. The primary goal of calibration is to continue training and to reinforce the scoring standards. Calibration sets may be "on the line" between score points or may contain unusual examples that are challenging

to score and therefore useful for reinforcing the scoring rubric. After scoring an online calibration set, scorers have an opportunity to ask questions of scoring supervisors and to seek clarification of the score point or annotation.

Scorer exception processing allows project managers to define intervals at which their scoring system checks scorer validity for exact and adjacent agreement. If scorers fall below preset standards, messages are automatically sent, interrupting their scoring process. Project leadership determines appropriate steps to remediate the scorer. The scorer may then work with a scoring supervisor, review anchor papers, or work though other activities to improve their scoring.

Through this process, the scoring system can automatically send an additional training/requalification set, and if performance is not improved, can lock scorers out of the scoring system. This automated process complements Pearson's supervisory methods and prevents scorers from continuing to score if standards are not maintained.

Validity responses are prescored responses strategically interspersed in the pool of live responses. These responses are not distinguishable from live responses and readers' scores are only accepted for monitoring purposes, not in replacement of the predetermined "true scores."

The validity mechanism provides an objective and systematic check of accuracy. It verifies that scorers are applying the same standards throughout the project and, therefore, guards against scorer drift and ultimately group drift. This procedure provides immediate feedback on individual scorers and the group as a whole.

The validity pool includes responses encompassing the entire score range for each item. Scorers will score these responses without being aware that they are validity responses, which will provide informative statistical scoring information. Validity responses will be sent to scorers throughout the project.

Select validity responses are annotated by the scoring director and flagged for review. If a reader scores one of these responses incorrectly, the scoring session is interrupted while the response appears on the scorer's screen with the true score, the score he or she assigned, and an annotation. This immediate feedback aids in preventing scorer drift before it occurs. Once a scorer has received feedback about a specific validity response, the response is flagged so the scorer does not receive it again.

Interpreting WritePlacer Results: WritePlacer Dimensions

In addition to the reported holistic score, feedback is provided on six dimensions considered essential in a well-written essay (College Board, 2018).

Purpose and Focus – The extent to which the writer presents information in a unified and coherent manner, clearly addressing the issue. Specific elements to consider include:

- Unity
- Consistency
- Coherence
- Relevance
- Audience

Organization and Structure – The extent to which the writer orders and connects ideas. Specific elements to consider include:

- Introduction
- Thesis
- Body paragraphs
- Transitions
- Conclusions

Development and Support – The extent to which the writer develops and supports ideas. Specific elements to consider include:

- Point of view
- Coherent arguments
- Evidence
- Elaboration

Sentence Variety and Style – The extent to which the writer crafts sentences and paragraphs demonstrating control of vocabulary, voice, and structure. Specific elements to consider include:

- Sentence length
- Sentence structure
- Usage
- Tone
- Vocabulary
- Voice

Mechanical conventions – The extent to which the writer expresses ideas using Standard English. Specific elements to consider include:

- Spelling
- Grammar
- Punctuation

Critical thinking – The extent to which the writer communicates a point of view and demonstrates reasoned relationships among ideas. Specific elements to consider include:

- Clarity
- Depth
- Precision
- Logic
- Accuracy
- Fairness
- Breadth
- Relevance

If dimension statements have been selected to be reported on the Individual Score Report, one of the dimension statements shown below in Table 5.1 will be reported for each of the indicated dimensions. Each statement describes the test taker's proficiency in the indicated dimension.

Table 5.1: *WritePlacer Dimensions and Descriptions*

Purpose and focus
Your response shows a clear purpose and a consistent focus.
Your response does not fully communicate purpose, and focus may be inconsistent.
Your response lacks clear purpose and focus.
Organization and structure
Your response demonstrates strong organization of ideas.
Your response demonstrates limited organization of ideas.
Your response demonstrates poor organization of ideas.

(Continued)

Development and support

Your response is logically developed and well supported.

Your response has limited support for your ideas.

Your response needs additional ideas and support.

Sentence variety and style

Your response shows skillful control of sentence structure and style.

Your response shows inconsistent control of sentence variety, word choice, and flow of thought.

Your response shows limited ability to vary sentence length and apply appropriate vocabulary.

Mechanical conventions

our response shows strong control of mechanical conventions
such as grammar, spelling, and punctuation.

Your response shows limited control of mechanical conventions
such as grammar, spelling, and punctuation.

Your response shows poor control of mechanical conventions
such as grammar, spelling, and punctuation.

Critical thinking

Your response shows clear and reasoned analysis of the issue.

Your response shows limited clarity and complexity of thought.

Your response shows insufficient reasoning and lacks complexity of thought.

5.2 Reporting

Score reporting

In keeping with the AERA/APA/NCME Standards, ACCUPLACER score reports have been developed at the student and institutional levels to provide their intended audiences with appropriate interpretations of the reports and guidelines outlining the appropriate use of test results. A variety of reports are available online 24/7 for all ACCUPLACER tests, including the following:

- Individual Score Report (ISR)
  - Generated for each student at the end of testing.
  - Shows student's identifying information and test scores, with CSEMs if option is selected by the institution.
  - Shows appropriated course placement, if placement rules have been entered into the testing site.
  - If WritePlacer is taken, the ISR includes the holistic score description and dimension statements.
- WritePlacer Response Report – Allows an institutional user to search and print essays submitted by students in response to a WritePlacer prompt.
- Placement Roster Report – Provides a list of students who placed into courses associated with a specific course group.
- Course Roster Report – Provides a list of students who placed into a specific course.
- Score Roster Report – Customizable report that may include students' scores, demographic information, and answers to background questionnaires as selected by the user for a specific date range.

Databases

ACCUPLACER data are stored in secure reporting databases and retained for five (5) years. All data are synched in real time with the Disaster Recovery environment so they will not be compromised during a disaster scenario. On a quarterly basis, the ACCUPLACER system removes testing data that are more than four years old. This routine maintenance of the amount of data stored in the system ensures that the platform provides immediate, stable, and accurate access to current student testing data.

5.3 Setting Placement Scores

To ensure valid course placement decisions using ACCUPLACER tests, appropriate placement (cut) scores must be established. Given that institutions differ greatly with respect to the composition of the student body and the faculty, as well as course content, it is not possible to stipulate specific test placement scores that should be used for placement decisions. Instead, each institution should establish their own placement scores to facilitate placement decisions based on factors and data unique to their institution. To support institutions in setting placement scores, College Board has prepared detailed guidelines for implementing the Bookmark standard setting method for ACCUPLACER placement tests. "A Guide to Conducting Standard Setting for the Next-Generation ACCUPLACER Placement Tests Using a Bookmark Procedure" is available through the link **https://accuplacer.collegeboard.org/pdf/guide-next-generation-standard-setting. pdf**. The document provides information for one way of setting placement scores, the Bookmark method (Mitzel, et al., 2001). This method is relatively easy to use and "perhaps the most popular method currently used to set performance standards on large-scale educational achievement tests" (Cizek, 2012, page 10), and has withstood legal challenges (see, for example, Lewis, et al., 1999 and Mitzel, et al., 2001). The document also provides guidance and information so that institutions are able to uphold AERA/APA/NCME Standard 5.21, which states that "the rationale and procedures used for establishing cut scores should be clearly documented" (AERA, APA, & NCME, 2014, p. 107).

5.4 ACCUPLACER Skills Insight and Proficiency Statements

Skills Insight and Proficiency Statements are designed to help students gain a better understanding of how scores relate to specific academic skills. They offer descriptions of performance and insight into skills measured at each score band. Skills Insight provides a set of data-driven statements intended to help students interpret their performance on the Reading, Writing, Arithmetic, QAS, and AAF. For ESL tests, Proficiency Statements help students interpret their performance. Skills Insight and Proficiency Statements are organized by test score band. Within each score band, the Skills Insight and Proficiency Statements describe what a student scoring within that band is likely to know and be able to do in relation to the academic skills measured on the tests. The goal of Skills Insight and Proficiency Statements is to help students, teachers, administrators, and other stakeholders understand what a test score means and, for scores below the highest range, how performance could be improved.

Determining Score Ranges for Skills Insight

Before Skills Insight statements could be developed, it had to be determined what score ranges (i.e., score bands) would be most effective in providing narrative information to students about the skills they most likely had mastered. These score ranges were determined by first reviewing the test score scale underlying the ACCUPLACER placement tests to ensure adequate coverage of the scale for the tests. Having too many score bands would not allow us to provide skills that were meaningfully different between score groups, and therefore, would not provide useful information to the students. Having too few score bands would not allow us to help pinpoint the strengths and weaknesses of the students represented in the different score bands. Regarding the length of each band, dividing the whole score range into equal parts was considered. In the end, the distribution of student scores across the scale was the main factor in the decision.

After analyzing all the data and reviewing different possible numbers and lengths of score bands, College Board concluded that the score ranges 236 and below, 237–249, 250–262, 263–275, and 276 and above would best distinguish students' skills so as to be helpful to students. These score ranges are based on the theoretical distribution of student performance on each placement test. Scaled scores 237, 250, and 263 (i.e., the lower bounds of the three middle ranges) represent the distribution quartiles. The lower bound of the highest range, scaled score of 276, represents the 90th percentile of the student performance distribution. Using these ranges, the Skills Insight statements for each of the first three ranges describe what 25% of students can be expected to know and be able to do. Furthermore, the Skills Insight for the highest range represents skills and knowledge that the top 10% of students can typically demonstrate.

Development of the Skills Insight Statements

Skills Insight statements for ACCUPLACER tests were developed using the item mapping methodology. By placing individual items along the score scale, item maps help illustrate what students at different score bands know and are able to do. Item descriptions focusing on the skills and knowledge required to respond correctly are used collectively to create Skills Insight statements.

Each item's position on the scale is determined by the likelihood of students responding to the item correctly. For the ACCUPLACER placement tests, items are placed on the scale using the response probability of 0.67. Using IRT, each item is placed or mapped to a score at which students have a 0.67 probability of selecting the correct response. Descriptions of items mapped to a selected band make up the statements of what the students whose score fall in that range know and are able to do.

5.5 Using Multiple Factors in Placement Decisions

Standard 12.10 of the APA/AERA/NCME Standards asserts that "In educational settings, a decision or characterization that will have major impact on a student should take into consideration not just scores from a single test but other relevant information." (AERA, APA, & NCME, 2014, p. 198) To provide guidance to institutions in using additional information when making placement decisions using ACCUPLACER placement tests, College Board released the paper Multiple Factors in College Placement Decisions, which is available through this link: **https://accuplacer.collegeboard.org/pdf/multiple-factors-college-placement-decisions.pdf**

Psychometrics

Once the ability scores on placement tests have been estimated through the CAT algorithm as described in Section 3.4, they must be transformed into a form that allows for interpretation and accounts for variance across test forms. This involves establishing a scale. The scale is a numerical system that conveys test performance. The first part of this chapter discusses the creation of this numerical system used to report results of the ACCUPLACER tests. This procedure is referred to as scaling. The second part of this chapter discusses the precision of the scale scores. The precision of the scale scores are then discussed.

6.1 Scaling

ACCUPLACER placement tests are computer-adaptive tests, using Item Response Theory (IRT) as the psychometric method for analyzing, calibrating, and choosing items administered to a test taker. The items chosen are based on a constantly updated evaluation of the test taker's ability on the material after each preceding question has been answered. With IRT, the computed score is an estimate of ability that ranges from -5.0 to 5.0. Since the -5.0 to 5.0 scale is difficult to interpret for test users, an alternative scale is used to report student performance on the test. According to AERA/APA/NCME Standard 5.2, "the procedures for constructing scales used for reporting scores and the rationale for these procedures should be described clearly" (AERA, APA, & NCME, 2014, p. 202). This section provides a walkthrough of the scaling procedures for ACCUPLACER placement tests.

Goals for the Scales

ACCUPLACER placement tests are on a 101-point scale. The Arithmetic; Quantitative Reasoning, Algebra, and Statistics (QAS); Advanced Algebra and Functions (AAF); Reading; and Writing Tests have scale scores ranging from 200 to 300, with a mean of 250. The English as Second Language (ESL) tests range from 20 to 120, with a mean of 70. Each ACCUPLACER placement score scale has a standard deviation of 20. The intended range and distributions of the scale scores were selected because they are deemed sufficient to support the interpretation of test performance for the purpose of placing students in the appropriate college courses.

Scaling Procedure

The scaling procedures for each ACCUPLACER placement test involve creating a scale score conversion table based on simulation results. Starting with the intended scale identified (e.g., range of 200 to 300, increment of 1) and the intended distribution decided (e.g., mean of 250 and standard deviation of 20), the next step is to determine the first two moments of the estimated abilities (i.e., θ) of the 10,000 simulees.

Let *SS=Scale Score* with

Range: SS $\in$[200, 300],

Mean: μ_{SS}=250,

and

Standard Deviation: σ_{SS}=20

and the first two moments of θ estimates are $\widehat{\mu_\theta}$ and $\widehat{\sigma_\theta}$.

The scaling formula is

$$SS=a+b\theta,$$

where

$$\mu_{SS} = a + b * \widehat{\mu_\theta}$$

and

$$\sigma_{SS} = b * \widehat{\sigma_\theta}.$$

From $\sigma_{SS} = b * \widehat{\sigma_\theta}$, compute the slope as

$$b = \frac{\sigma_{SS}}{\widehat{\sigma_\theta}}.$$

From $\mu_{SS} = a + b * \widehat{\mu_\theta}$, compute the intercept as

$$a = \mu_{SS} - b * \widehat{\mu_\theta}.$$

Using the resulting formula, a scale score is computed for each of the 10,000 simulees. The mean and standard deviation of the computed scale scores are determined. The values of and/or are adjusted and optimized as appropriate to make the computed scale score mean and standard deviation closer to 250 and 20, respectively. Using the resulting slope, a conversion table is constructed such that each scale score has a corresponding θ value. To satisfy a requirement of the ACCUPLACER platform, -5.0 is always mapped to 200 and 5.0 is always mapped to 300.

6.2 Reliability
What is Reliability?
Reliability is the degree to which scores arising from an assessment produce stable and consistent results. In other words, the reliability coefficient indicates the amount of consistency in scores. A score with a reliability coefficient of 1 is a perfectly reliable score, while a value of 0 means that the score is not at all reliable. For example, if a score has a reliability of, say 0.88, one may think about 0.88 as the amount of consistency.

Reliability and Conditional Standard Error of Measurement (CSEM) of Scale Scores
The standard error of measurement (SEM) provides an estimate of the amount of error in scores. SEM equals standard deviation times the square root of (1 minus test reliability). Based on the SEM formula, we can see that the SEM and reliability are inversely related. That is, the more reliable the score is, the smaller standard error of measurement the score has. In contrast, the less reliable the score is, the larger standard error of measurement the score has. The SEM is especially meaningful to a test taker because it applies to a single score and it uses the same units as the test.

For ACCUPLACER placement tests, a SEM is estimated for each score. That is, around each scaled score, a value is computed to indicate the level of certainty about where a student's true score may lie given the score that that student obtained. Each computed value represents the variability one would expect to see in the scaled scores of a test taker of a given ability who takes the test multiple times. This is referred to as the Conditional Standard Error of Measurement (CSEM). These values may be used to report a confidence interval within which a test taker's true score might fall, given that test taker's obtained score. For example, if a test taker receives a score of 250 on a placement test and the CSEM is 4.9, there is a 68% probability that the test taker's true score is within the 245.1 and 254.9 range. In other words, if that test taker took the test 100 times, and the range was computed each time, then approximately 68% of these ranges will contain the person's true score. A smaller value of CSEM provides more precise measurement.

The CSEMs for scale scores are estimated based on simulation results after the theta to scale score conversion table is developed. The following is a summary of the procedure to estimate the scale score CSEMs:

1. Create a uniform distribution of 50,000 true thetas ranging from -5 to +5 so that all theta levels are equally represented.
2. Use the true theta distribution generated in step 1 to simulate CAT so that each true theta gets a theta estimate.
3. Convert true thetas to true scale scores using the conversion table.
4. Convert estimated thetas to estimated scale scores using the conversion table.
5. At each integer true scale score point (e.g., 20–120 or 200–300), compute standard deviation of estimated scale scores as CSEM for the scale score. The unrounded scale score estimates are used in calculation for better precision.

The CSEM is an optional element for the individual student report (ISR) that is included at the discretion of the institution. The CSEM is also a very important piece of information for institutions when deciding on placement policies such as finalizing placement scores or using multiple factors in placement decisions. A new CSEM table is created whenever a CAT item pool is refreshed and is always current in the ACCUPLACER platform. The most current CSEM table is available upon request from College Board.

Interrater Consistency of WritePlacer Scores

To ensure that test takers' writing for WritePlacer is scored reliably, an annual audit of the automated scoring is implemented for all prompts. The result of this audit is summarized here.

For the 17 WritePlacer and five WritePlacer ESL prompts, Pearson conducts an annual audit using the following procedure. Five percent of all essays scored by the automated scoring engine throughout the year are selected for human scoring through the Pearson Performance Scoring Center. A Stratified Sampling Method is applied to select the essays that are included in the annual audit scoring. All essays are scored by a human scorer and machine scores and human scores are analyzed for exact and adjacent agreement interrater reliability (IRR). Resolution scoring is applied to nonadjacent essays if the original IRR results are below 90% on the prompt.

During the 2017 audit, the original IRR for three out of 17 WritePlacer prompts and three out of five WritePlacer ESL prompts were lower than 90% and thus needed resolution scoring for responses with nonadjacent scores. Added to the responses that needed resolution scoring for each prompt were an equal number of responses that had exact or adjacent scores selected randomly. Resolution scoring was performed by expert scorers only. For all the responses included in the resolution scoring, the expert scores replaced the original human scores. The result from this resolution process was that all 22 prompts exceeded the 90% exact or adjacent agreement expectation, as presented in Table 6.1. The overall IRR value was 94.6%.

Another index of interrater agreement commonly used in essay scoring is the quadratic weighted kappa. The value of this IRR index for each prompt is also included in Table 6.1.

To oversee the audit, College Board has a yearly review meeting with Pearson to go over the results of the WritePlacer audit. Members of the Psychometrics team participate in the annual review meeting, offering feedback for potential improvements that can be applied to future audits as appropriate.

Table 6.1: Interrater Reliability for WritePlacer and WritePlacer ESL Prompts

Prompt	Number of Essays Compared	Exact plus Adjacent Agreement	Quadratic Weighted Kappa
B11	1224	92.32%	0.74
B17	3070	98.01%	0.78
B24	2924	90.90%	0.71
B26	2936	91.04%	0.72
C04	3120	93.17%	0.72
C06	2575	91.92%	0.72
C14	1527	95.15%	0.76
C16	3065	93.12%	0.74
C22	1776	93.69%	0.69
C24	1606	90.78%	0.72
C28	1366	95.61%	0.74
C30	3392	93.60%	0.73
D02	3224	91.25%	0.70
D05	3064	93.67%	0.75
D12	1301	96.23%	0.77
D21	1451	93.45%	0.70
E01	1697	96.64%	0.76
ESL1	426	95.31%	0.73
ESL2	396	96.21%	0.77
ESL3	422	95.26%	0.74
ESL4	411	94.89%	0.75
ESL5	415	94.94%	0.71

CHAPTER 7
Validity

This final chapter covers validity. For the ACCUPLACER Suite, criterion-based evidence for validity will be collected and documented when the required data from tests, decisions, and course performance are available. In many ways, every chapter of this manual concerns itself with issues of validity, as all the procedures described in the previous sections attempt to ensure that ACCUPLACER tests produce scores that are valid assessments of the construct being tested. As such, a commitment to matters of validity is of paramount importance to us as we develop, administer, and score ACCUPLACER placement tests, as well as any other College Board assessment. We have placed a deeper examination of issues of validity here at the end of this manual, as we believe that by first gaining an understanding of the many processes involved in creating, administering, and scoring ACCUPLACER placement tests, and understanding how and why we interpret those scores for their intended uses, one can more completely comprehend the steps taken toward establishing sound validity evidence.

For ACCUPLACER, test validity must be evaluated with respect to the degree to which test takers' scores support making appropriate placement decisions. Arguably, all the evidence presented in the previous chapters goes toward supporting this claim, from the earliest stage of test development, right up to the final interpretation of scores. Our examination of validity as it relates to ACCUPLACER begins in the broadest terms, as Section 7.1 provides a brief overview of validity as a concept and the goals of test score validation. We then shift focus to the test itself. Section 7.2 presents the evidentiary foundations behind the test content found in ACCUPLACER placement tests.

7.1 Introduction to Validity as a Concept

Validity is not an intrinsic property of a test. Rather, it is the extent to which the inferences (interpretations) derived from test scores are justifiable from both scientific and equity perspectives. For decisions based on test scores to be valid, the use of a test for a particular purpose must be supported by theory and empirical evidence, and biases in the measurement process must be ruled out.

As many psychometricians have pointed out (e.g., Cronbach, 1971; Messick, 1989; Shepard, 1993), in judging the worth of a test, it is the inferences derived from the test scores that must be validated, not the test itself. Therefore, the specific purpose(s) for which test scores are being used must be considered when evaluating validity. For example, a test may be useful for one purpose, such as course placement, but not for another, such as college admission.

Contemporary definitions of validity in testing borrow largely from Messick, who stated that "validity is an integrated evaluative judgment of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of inferences and actions based on test scores or other modes of assessment" (Messick, 1989, p. 13). Based on this definition, validity is not something that can be established by a single study, and tests themselves cannot be labeled "valid" or "invalid." Given that validity is the most important consideration in evaluating the use of a test for a particular purpose, and such utility can never be unequivocally established, establishing that a test is appropriate for a particular purpose is an arduous task. Thus, the following facts about validity should be clear: (a) tests must be evaluated with respect to a particular purpose; (b) what needs to be validated are the inferences derived from test scores, not the test itself; (c) evaluating inferences made from test scores involves several different types of qualitative and quantitative evidence; and (d) evaluating the validity of inferences derived from test scores is not a one-time event, it is a continual process.

To make the task of validating inferences derived from test scores both scientifically sound and manageable, Kane (1992, 2006) proposed an "argument-based approach to validity." In this approach, the validator builds an argument based on empirical evidence to support the use of a test for a particular purpose. Although this validation framework acknowledges that validity can never be established absolutely, it requires evidence that the test measures what it claims to measure, that the test scores display adequate reliability, and that test scores display relationships with other variables in a manner congruent with the test's predicted properties. Kane's practical perspective is consistent with the Joint Standards (AERA, APA, & NCME, 2014), which provide detailed guidance regarding the types of evidence that should be brought forward to support the use of a test for a particular purpose. In keeping with the notion that all the statistical processes applied to a test aid in establishing validity, the 2014 AERA/APA/NCME Standards state that:

> *A sound validity argument integrates various strands of evidence into a coherent account of the degree to which existing evidence and theory support the intended interpretation of test scores for specific uses.… Ultimately, the validity of an intended interpretation … relies on all the available evidence relevant to the technical quality of a testing system. This includes evidence of careful test construction; adequate score reliability; appropriate test administration and scoring; accurate score scaling, equating, and standard setting; and careful attention to fairness for all examinees.… (AERA, APA, & NCME, 2014, p. 21-22).*

To build a validity argument for a test, there are several types of evidence that can be presented. Evidence based on content involves gathering data from content experts regarding the degree to which the behaviors sampled on the test represent the behaviors the test is designed to measure. Evidence based on criterion-related information involves evaluating correlations among test scores and other variables related to the construct measured. This evidence includes predictive and concurrent as special cases that involve correlating test scores with future or current criterion performance, respectively. Other evidence for the validity of interpreting test scores involves gathering data that show test scores are indicative of the construct measured. College Board is committed to continually building the validity argument for ACCUPLACER placement tests by continuing to collect validity evidence at every opportunity.

With each ACCUPLACER test being used by different institutions for placement to different courses, using different placement scores as part of different course placement policies, each institution is encouraged to evaluate the predictive placement validity of their placement decisions. To support institutions in this endeavor, College Board provides a free service through the Admitted Class Evaluation Service (ACES).

7.2 Content-Oriented Validity Evidence
What Does the ACCUPLACER Suite of Assessments Measure?
ACCUPLACER is intended to be used to assess the skills and knowledge that matter most for postsecondary education. Along with factors such as HSGPA, the resulting scores from the assessments are intended to be interpreted as indicators of a student's readiness for college and career training programs. According to *The Standards for Educational and Psychological Testing*, Standard 1.11, "When the rationale for test score interpretation for a given use rests in part on the appropriateness of test content, the procedures followed in specifying and generating test content should be described and justified with reference to the intended population to be tested and the construct the test is intended to measure or the domain it's intended to represent" (AERA, APA, & NCME, 2014, p. 26). This section describes the foundations for the decisions made about the content included in the ACCUPLACER Suite, which consist of the Reading, Writing, and Math placement tests; the English as a Second Language (ESL) tests; and the two essay tests in the Suite.

Content Validity for the ACCUPLACER Reading and Writing Placement Tests

The ACCUPLACER Reading Test is focused on the assessment of students' comprehension and reasoning skills in relation to appropriately challenging prose passages across a range of content areas. The Writing Test is focused on the assessment of students' revision and editing skills in the context of extended prose passages across a range of content areas.

A number of key design elements strongly supported by evidence are interwoven throughout the Reading and Writing Tests. These include:

- The use of a specified range of text complexity aligned to college and career readiness levels of reading
- A focus on words in context and on word choice
- Attention to source analysis and use of evidence
- Attention to a core set of important English language conventions and to effective written expression
- The requirement that students work with texts across a wide range of disciplines

Several of these elements are key features of the ACCUPLACER Reading and Writing Tests, as described in Chapter 3: Test Development Procedures and in extended discussions of test specifications (College Board, 2017c). Other elements, while not explicitly key features, are nonetheless important factors and have been incorporated into the assessment in a significant manner, in keeping with what research shows to be most essential for college readiness and success.

Text Complexity

Numerous studies have highlighted the long-standing gap between the high level of challenge posed by the required readings in college-entry, credit-bearing courses and workforce training programs and the comparatively simpler readings used in much of K–12 education, including many high school courses. For example, Adams (2009), reviewing the research literature on the challenges students face reading complex texts, helped collect a range of scholarly evidence documenting a several-decades-long decline in K–12 text complexity even as college and career readiness demands on students' reading skills remained high.

The ACCUPLACER Suite Reading and Writing Tests align the levels of text complexity represented in the tests' passages with the requirements of common first-year, credit-bearing college courses and workforce training programs. This alignment supports the emerging movement to close this preparedness gap by making text complexity a central part of the test design. Students taking the ACCUPLACER Suite Reading and Writing Tests are asked to engage with passages selected, in part, to exhibit a range of text complexities up through and including levels comparable to those expected of students entering college and workforce training programs. To ensure that texts on the ACCUPLACER Suite are appropriately complex—challenging but not inaccessible to college- and career-ready test takers—test development staff make use of feedback from secondary and postsecondary subject matter experts and test data on student performance as well as quantitative and qualitative measures of text complexity. Considered together, the ACCUPLACER Suite Reading and Writing Tests measure whether students can read, improve, and analyze texts at levels of difficulty required of incoming postsecondary students.

Words in Context

Research has shown the close link between students' vocabulary achievement and their success in reading and in school in general (Beck, McKeown, & Kucan, 2013). With a broad and deep vocabulary, readers are more likely to understand what they read and, in turn, to derive the meaning of words in the contexts in which they appear. Indeed, the role of vocabulary in reading comprehension is difficult to overstate, given the word richness

of text. A quick comparison between oral and written language indicates that while the conversation of college-educated adults contains an average of 17.3 rare words per thousand, even children's books exhibit almost double that frequency: 30.9 (Becker, 1977; Hayes & Ahrens, 1988; National Center for Education Statistics, 2013; National Reading Panel, 2000; Stanovich, 1986; Whipple, 1925).

Attaining skilled comprehension through vocabulary depends on how the vocabulary is acquired. Beck and her colleagues have sensibly focused on what they refer to as Tier Two words— "words that are of high utility for mature language users and are found across a variety of domains"—because they appear frequently in written texts (but uncommonly in oral language) across a wide range of subjects. (By contrast, Tier One words require little instruction for most students because they're generally acquired through conversation, and Tier Three words are either limited to a certain domain of knowledge—and thus are best studied as part of work in that domain—or too rare to be found with any frequency in written text.) Other researchers have reached a similar conclusion about the need to concentrate instruction on high-utility words (Beck et al., 2013; Nation, 2001; Stahl & Nagy, 2006).

There is a sharp focus on vocabulary in the Reading and Writing Tests. In the Reading Test, students are called on to determine the meaning of vocabulary in context, with an emphasis on Tier Two words and phrases. In the Reading and Writing Tests, students are also presented with other vocabulary-related challenges, including analyzing word choice rhetorically and improving the precision, concision, and context appropriateness of expression.

Source Analysis and Evidence Use

Students' developed abilities to analyze source texts and, more broadly, to understand and make effective use of evidence in reading and writing are widely recognized as central to college and career readiness and success. National curriculum surveys conducted by College Board and others demonstrate that postsecondary instructors rate high in importance such capacities as summarizing a text's central argument or main idea, identifying rhetorical strategies used in a text, and recognizing logical flaws in an author's argument, as well as writing analyses and evaluations of texts, using supporting details and examples, and developing a logical argument (Achieve, Inc., The Education Trust, & Thomas B. Fordham Foundation, 2004; ACT, Inc., 2009; College Board, 2006; Kim, Wiley, & Packman, 2012; Seburn, Frain, & Conley, 2013). Institutions such as Duke University, Cornell University, Texas A&M University, and the University of California, Berkeley, have devoted considerable resources to developing the skills of source analysis and evidence use in their students.

The Reading and Writing Tests support an emphasis on source analysis and evidence use throughout. The Reading Test requires students to answer questions based on what is stated and implied in texts across a range of content areas. The Writing Test includes questions asking students to develop, support, and refine claims and ideas in multiparagraph passages and to add, revise, or delete information in accordance with rhetorical purpose.

Language Conventions and Effective Language Use

In addition to vocabulary knowledge and use, skilled expression in language includes understanding and observing the conventions of standard written English and, more generally, making informed, thoughtful grammatical choices. Knowledge of conventions includes learning and adhering to language "rules" governing conventional expression as well as knowledge of the practices that lend precision and clarity to writing, aid comprehension, and facilitate academic success. Language practices have been described in terms of grammatical choices as representing the relationships between writers and their world and as expressions of how writers attend to the words of others and position themselves in relation to others (Micciche, 2004).

The Reading and Writing Tests support a thoughtful emphasis on language conventions and language use in several important ways. Effective language use and mastery of a core set of conventions linked with college and career readiness and success are two key elements of the Writing Test, which, among other aims, assesses students' application of these skills in the context of high-quality multiparagraph passages that must be revised and edited. The Reading Test includes questions that address students' capacity to analyze word choice rhetorically.

Disciplinary Literacy

Shanahan, Shanahan, and Misischia (2011) are prominent among those who have made the case in recent years that students' literacy development shouldn't be seen as merely the development of generic communication skills but instead should be grounded in making students familiar with the differing literacy demands of particular fields of study. These authors claim that reading, for example, is an importantly different activity when it's done in, say, a history, a math, or a chemistry context: "In addition to the 'domain knowledge' of the disciplines . . . each discipline possesses specialized genre, vocabulary, traditions of communication, and standards of quality and precision, and each requires specific kinds of reading and writing to an extent greater than has been recognized by teachers or teacher preparation programs" (Shanahan et al., 2011, p. 395). For example, instructors of entry-level health courses surveyed in California rated the knowledge of appropriate terminology in the healthcare setting as being very important for postsecondary success in the health sciences and medical technology career cluster (McGaughy et al., 2012).

The ACCUPLACER Reading and Writing Tests support a strong emphasis on disciplinary literacy through careful passage selection and question development. In the Reading and Writing Tests, students are expected to engage with and analyze appropriately challenging texts spanning numerous content areas, including literature and literary nonfiction, history/social studies, the humanities, science, and careers-related topics. Moreover, while questions on the Reading and the Writing Tests don't require students to have prior knowledge of specific topics in the content areas, these questions do, where possible and beneficial, reflect differences in the ways different disciplines approach literacy. Reading questions relating to a literature selection, for example, might address theme, mood, figurative language, or characterization—concepts that are generally not relevant to the sciences. Reading questions relating to a science selection, on the other hand, might require students to delineate the experimental process described in a text, analyze research data, or determine which conclusion is best supported by a study's findings—skills generally not required to comprehend literary texts.

Content Validity for the ACCUPLACER Suite Math Tests

The overall aim of the Math Tests is to assess students' fluency with, understanding of, and ability to apply the mathematical concepts, skills, and practices that are most strongly prerequisite for and useful across a range of college majors and careers, as defined by their grade level.

As with Reading and Writing, a number of key design elements strongly supported by evidence are interwoven through the Math area. Evidentiary foundations for the Math Tests include:

- A focus on content that matters most for college and career readiness
- An emphasis on problem solving and data analysis
- The inclusion of both calculator and no calculator questions as well as attention to the use of a calculator as a tool

Several of these elements are key features of the Math Tests, as described in Section 3.1 and in extended discussions of test specifications (College Board, 2017c). Other elements, while not necessarily key features, are nonetheless important factors and have been incorporated into the assessment in a significant manner, in keeping with what research shows to be most essential for college readiness.

Focusing on Content That Matters Most

There is a major disconnect today in math between the K–12 and higher education systems. In a recent national survey, high school teachers and postsecondary instructors were asked whether students were leaving high school very well prepared for college-level math. While 37% of high school teachers said yes, only 4% of postsecondary instructors agreed (Sanoff, 2006).

Surveys of postsecondary faculty and studies of entry-level postsecondary course demands have repeatedly pointed to the conclusion that postsecondary instructors value greater command of a smaller set of prerequisites over shallow exposure to a wide array of topics. As one survey noted, because the postsecondary survey results indicate that a more challenging treatment of fundamental content knowledge and skills needed for credit-bearing college courses would better prepare students for postsecondary school and work, states would likely benefit from examining their state standards and, where necessary, refocusing them on the knowledge and skills that research shows are essential to college and career readiness and postsecondary success (ACT, Inc., 2009).

In October 2013, the Council of Chief State School Officers released a set of summative assessment principles for ELA/literacy and math assessments aligned to college and career readiness standards. These assessment principles are meant to form the basis for states' evaluations of their assessment systems. The principles greatly stress the importance of focusing summative assessments on what matters most. The very first alignment principle in math is that of "focusing strongly on the content most needed for success in later mathematics." As the document notes, "In a [college- and career-ready] aligned assessment system . . . high school focuses on widely applicable prerequisites for careers and postsecondary education" (Council of Chief State School Officers, 2013, p. 2).

One of the most important ways the ACCUPLACER Math Tests (Arithmetic; Quantitative Reasoning, Algebra, and Statistics; and Advanced Algebra and Functions) address the gap between postsecondary and K–12 expectations is through the assessment's concentrated focus on the content that matters most for postsecondary education. In a national survey published in 2011, Conley reinforced the conclusion that some content areas require much stronger emphasis than others. The distinctive importance of algebra is unmistakable from Conley's data. Other math domains have a more mixed profile, typically including more material that isn't as relevant to most postsecondary work and/or isn't a prerequisite for most postsecondary work. The data from this study directly support the content choices made in the Math Tests (Conley et al., 2011).

Problem Solving and Data Analysis

There is ample evidence that problem solving and data analysis—the ability to create a representation of a problem, consider the units involved, attend to the meaning of quantities, and know and use different properties of operations and objects—are important. Quantitative literacy is part of participation in a democracy; it's important to employers, who need students who can use math outside of the classroom; and it's important not only for science, technology, engineering, and math (STEM) fields but also for a wide range of college majors (Conley, 2006; Conley, McGaughy, Brown, van der Valk, & Young, 2009; National Council on Education and the Disciplines, 2001).

A recent study by the National Center on Education and the Economy (2013) that analyzed the actual mathematical demands of course syllabi and assignments in two-year institutions also supports the emphasis of the ACCUPLACER Math Tests on problem solving and data analysis. The study found that students pursuing two-year degree programs must be able to work with multistep problems involving ratios, proportional relationships, percentages, unit conversions, and complex measurement problems.

Such problems are an ideal connection point for science and for college and career readiness because so many of the quantities in applied science involve proportional relationships and/or are formed by division (such as rates, densities, and gradients). The Problem Solving and Data Analysis questions on the Math Tests contain multipart problems.

Calculator and No Calculator Questions

A calculator is a tool, and decisions about when and when not to use it involve a variety of considerations. The data are clear that postsecondary instructors expect students to be fluent in rational number arithmetic (ACT, Inc., 2007, 2009). Including the no calculator sections on the ACCUPLACER Math Tests helps assure postsecondary instructors that students who earn high scores on the tests don't lack the basic prerequisites. Questions with calculator options are designed to probe students' ability to make wise choices between when to use the calculation and when not to use it. For some questions, the calculator lends efficiency; for others, the ability to make use of structure or to reason abstractly leads to the most rapid solution.

Content Validity for the ACCUPLACER Suite ESL Tests

The overall aim of the English as a Second Language (ESL) tests in the ACCUPLACER Suite is to assess English Language Learners' (ELL) command of core English language competencies. Scores from these tests help determine which level of language instruction would be most appropriate to ensure that these students receive the support they need so they're positioned for successful acquisition of subject matter content in postsecondary courses as well as effective communication in and out of the classroom.

ESL placement decisions can have a great impact on students' academic lives and potentially their future careers. For example, being placed into developmental ESL classes lengthens the time needed to complete the degree and increases a student's financial burden. It's thus critically important that tests used to determine ELLs' ESL placement are valid for this use and represent all facets of the constructs assessed. The ACCUPLACER ESL tests focus on questions that assess ELLs' abilities in four skills areas central to a successful path to college and career. In ACCUPLACER ESL Reading Skills, Sentence Meaning, Language Use, and Listening, students are tested on their developed ability to comprehend sentences, passages, and spoken discourse as well as their developed ability to use correct forms and grammatical conventions. The developed ability to fulfill tasks in these tests—whether assessed through questions requiring the understanding and command of gist, supporting details, rhetoric, alternative ways of conveying meaning, or grammatical, lexical, and syntactic forms—is measured using reading and listening passages simulating language used in both academic disciplines and everyday situations and interactions.

As with other tests in the ACCUPLACER Suite, a number of key design elements are interwoven throughout the four multiple-choice ESL tests. Among these are:

- Attention to a core set of knowledge and skills ELLs need for effective learning and communication
- The requirement that students work with texts across a wide range of disciplines
- An emphasis on content and contexts accessible to ELLs
- An emphasis on reflecting the diversity of ELLs

Several of these elements are key features of the ESL Tests, as described in Chapter 3: Test Development Procedures. Other elements, while not necessarily key features, are nonetheless important factors and have been incorporated into the assessment in a significant manner, in keeping with what research shows to be most essential for college readiness and success.

Knowledge and Skills That Matter Most

The ESL tests focus on a core set of English language skills and conventions that ELLs need to be proficient in if they're to successfully acquire content knowledge in college courses and workforce training programs, communicate effectively, and be positioned well for learning further language skills. These include the ability to comprehend meaning—both literal and implied—in reading and listening passages; infer purpose, tone, gist, and meaning of unfamiliar words from context; locate supporting details;

draw logical conclusions; and observe standard English conventions in texts on topics that students would encounter in postsecondary courses, college environments, and workforce training programs.

Range of Disciplines

Many students who have demonstrated proficiency for gaining admission to college experience language-related challenges in their coursework (Chang & Kanno, 2010). It's thus imperative that ESL tests that place ELLs into academic courses cover—to the extent possible and without introducing construct-irrelevant challenges—language of the discipline that students must learn and use to participate and engage in meaningful ways in the content area. Selected texts and tasks aimed at addressing such language demands include those that assess the ability to understand common discipline-specific words and phrases using context clues; texts that use academic register; and language important to specific disciplines (e.g., in arts/humanities, language for describing and comparing/contrasting). Care is also taken to ensure selected texts cover a meaningful range of casual and academic contexts and represent modes ELLs need to be familiar with, such as narrative, informative/explanatory, and, to a lesser extent, argumentative.

Fair and Accessible Contexts

As ACCUPLACER ESL tests must allow the classification of ELL students from a wide range of backgrounds, experiences, and degrees of knowledge of English and English-speaking culture into levels for delivering appropriate instruction, there is necessarily a focus on fairness and accessibility; test takers must be able to understand the texts, questions, and tasks included so they have the opportunity to demonstrate what they know and can do. To the extent possible, these tests must be free of elements that interfere with test takers' opportunity to demonstrate learned abilities. Factors that research has shown to be impediments include sentence structures that may be confusing or difficult to follow, such as passive voice or sentences with multiple clauses (Forster & Olbrei, 1973; Schachter, 1983); cultural references or idiomatic expressions that may be unfamiliar (Bernhardt, 2005); syntax that may be confusing or ambiguous, such as double negatives (Cummins, Kintsch, Reusser, & Weimer, 1988); challenging vocabulary or overly complex words that aren't related to the construct assessed (Abedi, 2006; Bailey, Huang, Shin, Farnsworth, & Butler, 2007). Therefore, every effort is made to ensure test content and contexts are accessible, including using simplified language where appropriate and reducing unnecessary language or information.

Diversity

As the ESL tests are taken by students of diverse backgrounds, the question pool includes substantial content that visibly reflects U.S.-based racial and ethnic diversity and international or global perspectives, cultures, and settings. Stimuli and questions also reflect a balanced representation of genders (including passages that focus on women and girls and their experiences).

Content Validity for the ACCUPLACER Essay Tests

WritePlacer and WritePlacer ESL are the two ACCUPLACER essay tests. WritePlacer is designed to help postsecondary administrators assess students' readiness to successfully meet the writing demands of credit-bearing courses, while WritePlacer ESL helps colleges determine the appropriate level of English instruction to provide ELLs either before or as they take college-level courses. Both focus on the assessment of test takers' developed ability to successfully write an original essay in which they develop a point of view or position in response to a writing task using reasoning, personal experience, observations, and an appropriate rhetorical approach. As performance assessments, these tests may offer opportunities for test takers to present a more comprehensive picture of what they know and are able to do than multiple-choice tests

are able to do. For WritePlacer and WritePlacer ESL to yield meaningful outcomes, and because test scores are used to make important decisions about students (whether it's to place them into college or ESL courses), it's critical that:

- Each writing prompt successfully elicit a performance of writing
- The scoring criteria, or rubrics, clearly articulate the definitions of the construct measured (i.e., writing ability in the context of college students and college-bound ELLs, respectively), and
- The *Intelligent Essay Assessor (IEA) that electronically scores the essays be trained to interpret these scoring criteria*

This section discusses the first two imperatives above, while information on scoring and the IEA is available in Chapter 5: Interpretation of Results.

Research suggests that some variables unrelated to the focal measurement construct (e.g., unnecessary linguistic complexity, cultural biases in construction of prompts) can affect the trustworthiness of test scores (Abedi, 2006; Solano-Flores & Trumbull, 2003; Solano-Flores, 2008), thus negatively impacting an assessment's usefulness as a tool for evaluating student learning and informing instruction. To make sure the writing tasks (i.e., prompts) and criteria for evaluating responses elicited by these are relevant to the construct and intended score interpretation, College Board works closely with writing experts, including writing faculty members from high schools, two-year colleges, and four-year colleges from around the country, to define and operationalize the construct, review and refine prompts, and review scoring criteria and field test data. Additionally, an audit of IEA scoring is undertaken annually.

To ensure that WritePlacer and WritePlacer ESL prompts elicit the components of writing that are included in the definition of the construct assessed—no more and no less, as to do the former would result in construct-irrelevant variance and to do the latter would lead to underrepresentation of the construct—several key design elements are interwoven into both essay tests. These include:

- Accessible prompts that allow test takers to respond in a variety of ways
- A focus on core writing skills required of most college courses.

Several of these elements are key features of WritePlacer and WritePlacer ESL, as described in Chapter 3: Test Development Procedures. Other elements, while not necessarily key features, are nonetheless important factors and have been incorporated into the assessment in a significant manner, in keeping with what research shows to be most essential for college readiness and success.

Accessible Prompts

As all students should have an opportunity to demonstrate their knowledge and skills in the construct being assessed, WritePlacer and WritePlacer ESL prompts emphasize accessibility in relation to the writing task. WritePlacer and WritePlacer prompts are designed to stimulate critical thinking and to allow test takers to fulfill the writing task by drawing on a wide range of ideas and personal experiences. Prior topic-specific knowledge, including knowledge of U.S. culture or norms, isn't tested. Prompts are also relatively short (up to 80 words for WritePlacer and one to three sentence long for WritePlacer ESL). Issues and topics posed in the prompts are relevant to any number of fields and are stated in a straightforward manner; challenging language, vocabulary, and complex structures are minimized.

Skills That Matter Most

WritePlacer and WritePlacer ESL focus on a core set of skills students need in order to be proficient writers. In WritePlacer, the six essential skills assessed using the rubrics used for scoring the essays are purpose and focus (or the extent to which the writer presents information in a unified and coherent manner); organization and structure (the extent to which the writer orders and connects ideas); development and support (the extent to which

the writer develops and supports ideas); sentence variety and style (the extent to which the writer crafts sentences and paragraphs demonstrating control of vocabulary, voice, and structure); mechanical conventions (the extent to which the writer expresses ideas using standard English conventions); and critical thinking (the extent to which the writer communicates a point of view and demonstrates reasoned relationships among ideas).

WritePlacer ESL focuses on four writing characteristics considered critical for the target test-taking population to have a firm command of: word use (the extent to which the writer is able to use a wide range of words and phrases accurately); sentence use (the extent to which the writer is able to use a variety of sentence patterns with both independent and dependent clauses); grammar (the extent to which the writer is able to express ideas using grammatically correct English); and organization and development (the extent to which the writer is able to focus on the assigned topic and to develop ideas clearly).

References

Abedi, J. (2006). Language issues in item-development. In S. M. Downing & T. M. Haladyna (Eds.), *Handbook of test development* (pp. 377–398). Mahwah, NJ: Erlbaum.

Achieve, Inc., The Education Trust, and Thomas B. Fordham Foundation. (2004). The American Diploma Project: Ready or Not: Creating a High School Diploma that Counts. Washington, DC: Achieve, Inc.

ACT, Inc. (2007). ACT National Curriculum Survey 2005–2006. Iowa City, IA: Author.

ACT, Inc. (2009). ACT National Curriculum Survey 2009. Iowa City, IA: Author.

Adams, M.J. (2009). The Challenge of Advanced Texts: The Interdependence of Reading and Learning. In E. H. Hiebert (Ed.). Reading More, Reading Better: Are American Students Reading Enough of the Right Stuff? New York: Guilford.

AERA, APA, & NCME. (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

Bailey, A. L., Huang, B. H., Shin, H W., Farnsworth, T., & Butler, F. A. (2007). *Developing academic English language proficiency prototypes for 5th grade reading: Psychometric and linguistic profiles of tasks* (CSE Technical Report 727). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

Baker, F. B. & Kim, S. H. (2004). *Item Response Theory: Parameter Estimation Techniques.* Monticello, NY: Marcel Dekker, Inc.

Barry, C. L., & Niu, S. (2013). *Examining the impact of state-driven AP enrollment policies on student access to rigorous course work* (College Board Research Note 2013-4). New York, NY: College Board.

Beck, I.L., McKeown, M.G., & Kucan, L. (2013). Bringing Words to Life: Robust Vocabulary Instruction (2nd Ed). New York: Guilford Press.

Becker, W.C. (1977). Teaching Reading and Language to the Disadvantaged — What We Have Learned from Field Research. Harvard Educational Review. 47 (4), 518-543.

Bernhardt, E. (2005). Progress and procrastination in second language reading. *Annual Review of Applied Linguistics,* 25, 133–150.

Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord & M. R. Novick (Eds.). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.

Bock, R. D., & Mislevy, R. J. (1982). Adaptive EAP estimation of ability in a microcomputer environment. *Applied Psychological Measurement*, *6*, 431–444.

Cai, L. (2017). FlexMIRT: *Flexible multilevel multidimensional item analysis and test scoring* (version 3.51) [Computer software]. Chapel Hill, NC: Vector Psychometric Group.

Camilli G., & Shepard, L.A. (1993). *Methods for identifying biased test items.* Thousand Oaks, CA: Sage Publications.

Chang, Y. J., & Kanno, Y. (2010). NNES doctoral students in English-Speaking academy: The nexus between language and discipline. Applied Linguistics, 31(5), 671-692.

Cizek, G. J. (Ed.) (2012). *Setting performance standards: Foundations, methods, and innovations.* New York: Routledge.

Clauser, B. E. & Mazor, K. M. (1998). Using Statistical Procedures to Identify Differentially Functioning Test Items. *Educational Measurement: Issues and Practice, 17(1), 31-44.*

College Board. (2006). *College Board standards for college success: English language arts.* New York: Author. Retrieved from: **http://www.asainstitute.org/conference2008/featuredsessions/collegeboard-english-language-arts_cbscs.pdf**

College Board. (2011). *Guidelines on the Uses of College Board Test Scores and Related Data.* New York, NY: College Board.

College Board. (2014). The redesigned SAT: Evidentiary foundation. In *Test specifications for the redesigned SAT* (pp. 24–38). New York, NY: Author.

College Board (2017a). *ACCUPLACER COMPANION Administrator's Manual.* New York, NY: College Board. Retrieved from **https://accuplacer.collegeboard.org/pdf/accuplacer-companion-manual.pdf**

College Board (2017b). *ACCUPLACER User's Guide.* New York, NY: College Board. Retrieved from **https://accuplacer.collegeboard.org/pdf/accuplacer-user-guide.pdf**

College Board (2017c). Next-Generation ACCUPLACER Test Specifications, Version 2.0. New York, NY: College Board. Retrieved from **https://accuplacer.collegeboard.org/sites/default/files/next-generation-test-specifications-manual.pdf**

College Board (2018). ACCUPLACER Program Manual. New York, NY: College Board. Retrieved from **https://secure-media.collegeboard.org/digitalServices/pdf/accuplacer/accuplacer-program-manual.pdf**

Conley, D.T. (2006). College Board Advanced Placement® Best Practices Course Study. Eugene, OR: Educational Policy Improvement Center.

Conley, D.T., Drummond, K.V., de Gonzalez, A., Rooseboom, J., & Stout, O. (2011). Reaching the Goal: The Applicability and Importance of the Common Core State Standards to College and Career Readiness Eugene, OR: Educational Policy Improvement Center.

Conley, D.T., McGaughy, C., Brown, D., van der Valk, A., & Young, B. (2009). Validation Study III: Alignment of the Texas College and Career Readiness Standards with Courses in Two Career Pathways. Eugene, OR: Educational Policy Improvement Center.

Council of Chief State School Officers. (2013). States' Commitment to High-Quality Assessments Aligned to College- and Career-Readiness. Washington, DC: Author.

Cox, D.R. and Snell, E.J. (1989). *Analysis of binary data*. (2nd ed.). Chapman & Hall.

Cronbach, L.J. (1971). Test validation. In R.L. Thorndike & Angoff, W.G. (Eds.), *Educational measurement* (2nd ed., pp. 443–507). Washington, DC: American Council on Education.

Cummins, D. D., Kintsch, W., Reusser, K., & Weimer, R. (1988). The role of understanding in solving word problems. *Cognitive Psychology*, 20, 405–438.

Fan, M. (2007). *A comparison of automated test assembly programs for constructing parallel test forms-new and old*. Paper presented at the annual meeting of the Psychometric Society, Tokyo, Japan.

Forster, K. I. & Olbrei, I. (1973). Semantic heuristics and syntactic trial. *Cognition*, 2, 319–347.

Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory: principles and applications*. Boston, MA: Kluwer.

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. Newbury Park, CA: Sage Publications.

Hayes, D.P., & Ahrens, M.G. (1988). Vocabulary Simplification for Children: A Special Case of 'Motherese'? Journal of Child Language. 15, 395–410.

Holland, P. W., & Wainer, H. (Eds.). (1993). *Differential item functioning*. Hillsdale NJ: Erlbaum.

Jodoin, M. G., & Gierl, M. J. (2001). Evaluating type I error and power rates using an effect size measure with the logistic regression procedure for DIF detection. *Applied Measurement in Education, 14*, 329–349.

Kane, M. T. (1992). An argument based approach to validity. Psychological Bulletin, 112, 527-535.

Kane, M. T. (2006). Validation. In R. L. Brennan (Ed.), Educational measurement (4th ed., pp. 17–64). Washington, DC: The National Council on Measurement in Education & the American Council on Education.

Kim, Y., Wiley, A., & Packman, S. (2012). National Curriculum Survey on English and Mathematics. New York: The College Board.

Kingsbury, G. G., & Zara, A. R. (1989). Procedures for selecting items for computerized adaptive tests. Applied Measurement in Education, 2, 359–375.

Lewis, D.M., Green, D.R., Mitzel, H.C., Baum, K., & Patz, R.J. (1999). The bookmark standard setting procedure: Methodology and recent implications. Monterey, CA: McGraw-Hill.

Lord F.M. (1980) Applications of Item Response Theory to Practical Testing Problems. Hillsdale NJ: Erlbaum.

Lord, F. M., & Novick, M. R. (1968). Statistical theories of mental test scores. Reading, MA: Addison-Wesley.

McGaughy, C., Bryck, R., & de González, A. (2012). California Diploma Project Technical Report III: Validity Study; Validity Study of the Health Sciences and Medical Technology Standards. Eugene, OR: Educational Policy Improvement Center.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), Educational measurement (3rd ed., pp. 13–103). New York: American Council on Education and Macmillan.

Micciche, L. R. (2004). Making a case for rhetorical grammar.

Mitzel, H. C., Lewis, D. M., Patz, R. J., & Green, D. R. (2001). The bookmark procedure: Psychological perspectives. In G. J. Cizek (Ed.), Setting performance standards: Concepts, methods, and perspectives (pp. 249–281). Mahwah, NJ: L. Erlbaum.

Nation, I. S. P. (2001). Learning Vocabulary in Another Language. New York: Cambridge University Press.

National Center for Education Statistics. (2013). The Nation's Report Card: Vocabulary Results from the 2009 and 2011 NAEP Reading Assessments (NCES 2013–452). Washington, DC: Institution of Education Sciences, U.S. Department of Education.

National Center on Education and the Economy. (2013). What Does It Really Mean to Be College and Work Ready? The Mathematics Required of First Year Community College Students. Washington, DC: Author.

National Council on Education and the Disciplines. (2001) Mathematics and Democracy: The Case for Quantitative Literacy. Princeton, NJ: Author.

National Reading Panel (U.S), National Institute of Child Health and Human Development. (2000). Report of the National Reading Panel: Teaching Children to Read; An Evidence-Based Assessment of the Scientific Research Literature on Reading and Its Implications for Reading Instruction; Reports of the Subgroups. Washington, DC: National Institute of Child Health and Human Development.

Roderick, M., Coca, V., & Nagaoka, J. (2011). High school effects in shaping urban students' participation in college application, choice, and enrollment. Sociology of Education, 84(3), 178–211.

Sanoff, A.P (2006). What professors and teachers think: A perception gap over students' preparation. Chronicle of Higher Education. March 10, 2006. Vol. 52, Issue 27, p. B9.

Schachter, P. (1983). On syntactic categories. Bloomington, IN: Indiana University Linguistics Club.

Seburn, M., Frain, S., and Conley, D.T. (2013). Job Training Programs Curriculum Study. Eugene, OR: Educational Policy Improvement Center.

Segall, D. O. & Davey, T. (1995). Some new methods for content balancing adaptive test. Paper presented at the Annual Meeting of the Psychometric Society, Minneapolis, MN.

Shanahan, C., Shanahan, T., and Misischia, C. (2011). Analysis of Expert Readers in Three Disciplines: History, Mathematics, and Chemistry. Journal of Literacy Research. 43 (4), 393-429.

Shepard, L. A. (1993). Evaluating test validity. Review of Research in Education, 19, 405-450.

Sireci, S. G. (2001). *Analysis of Differential Item Functioning Across Females and Males on the Levels of English Proficiency Tests*. Report prepared for the College Board.

Solano-Flores, G. (2008). Who is given tests in what language by whom, when, and where? The need for probabilistic views of language in the testing of English language learners. *Educational Researcher*, 37(4), 189–199.

Solano-Flores, G., & Trumbull, E. (2003). Examining language in context: The need for new research and practice paradigms in the testing of English-language learners. *Educational Researcher,* 32(2), 3–13.

Stahl, S.A. & Nagy, W.E. (2006). Teaching Word Meanings. Mahwah, NJ: Erlbaum.

Stanovich, K.E. (1986). Matthew Effects in Reading: Some Consequences of Individual Differences in the Acquisition of Literacy. Reading Research Quarterly. 21, 360-406.

Swaminathan, H., & Rogers, H. J. (1990). Detecting differential item functioning using logistic regression procedures. *Journal of Educational Measurement*, *27*, 361-370.

Veldkamp, B. P., & van der Linden, W. J. (2000). Designing item pools for computerized adaptive testing. In W. J. van der Linden and C. A. W. Glas (Eds.), *Computerized adaptive testing: Theory and Practice.* Boston, MA: Kluwer Academic Publishers.

Wainer, H. (2000). *Computerized-adaptive testing: A primer (2nd ed).* Mahwah, NJ: Erlbaum.

Whipple, G.M. (1925). Report of the National Committee on Reading: Twenty-Fourth Yearbook of the National Society for the Study of Education, Part 1. Bloomington, IN: Public School Publishing Company.

Wyatt, J., Kobrin, J., Wiley, A., Camara, W. J., & Proestler, N. (2011). *SAT benchmarks: Development of a college readiness benchmark and its relationship to secondary and postsecondary school performance* (College Board Research Report No. 2011-5). New York, NY: College Board.

Wyatt, J., Wiley, A., Camara, W. J., & Proestler, N. (2011) The Development of an Index of Academic Rigor for College Readiness. College Board Research Report #2011-11. New York, NY: The College Board.

Zimowski, M. F., Muraki, E., Mislevy, R. J., & Bock, R. D. (1996). *BILOG-MG 3: Multiple-group IRT analysis and test maintenance for binary items* [computer software]. Chicago: Scientific Software International.

Zumbo, B. D. (1999) *A Handbook on the Theory and Methods of Differential Item Functioning (DIF): Logistic Regression Modeling as a Unitary Framework for Binary and Likert-Type (Ordinal) Item Scores.* Ottawa, ON: Directorate of Human Resources Research and Evaluation, Department of National Defense.