Wisconsin ACT® Technical Manual

Prepared by: WDPI & ACT

Spring 2025

Version 1.0

Preface

The purpose of this manual is to provide the state of Wisconsin and their stakeholders with technical information regarding the ACT assessment. Additional information can be found in *The ACT® Technical Manual*, which provides technical information about the ACT assessment, including national-level reliability, scaling and equating, and validity evidence. This technical report provides Wisconsin-specific information based on the 2024–2025 academic year.

The principal purpose of *The ACT® Technical Manual* (ACT, 2025) is to document technical characteristics of the ACT® test in light of its intended uses and interpretations. *The ACT Technical Manual* documents the collection of validity evidence that supports appropriate interpretations of test scores and describes various content-related and psychometric aspects of the ACT. Multiple test design and development processes are articulated documenting how ACT builds the assessment in line with the validity argument and how concepts like construct validity, fairness, and accessibility are attended to throughout the process. Also described are routine analyses designed to support continuous improvement and research intended to ensure that the program remains both psychometrically and educationally sound.

We encourage individuals who want more detailed information on a topic discussed in this manual, or on a related topic, to contact Wisconsin Department of Public Instruction (WDPI) or ACT.

Commitment to Fair Testing

ACT endorses and is committed to complying with *The Standards for Educational and Psychological Testing* (American Educational Research Association, American Psychological Association & National Committee on Measurement in Education, 2014). ACT also endorses the *Code of Fair Testing Practices in Education* (Joint Committee on Testing Practices, 2004), which is a statement of the obligations to test takers of those who develop, administer, or use educational tests and test data in the following four areas: developing and selecting appropriate tests, administering and scoring tests, reporting and interpreting test results, and communicating with test takers. ACT endorses and is committed to complying with the *Code of Professional Responsibilities in Educational Measurement* (NCME Ad Hoc Committee on the Development of a Code of Ethics, 1995), which is a statement of professional responsibilities for those involved with various aspects of assessments, including development, marketing, interpretation, and use.

We encourage individuals who want more detailed information on a topic discussed in this manual, or on a related topic, to contact ACT.

Please direct comments or inquiries to the address below:

Research Services
ACT Education Corp.
500 ACT Drive
Iowa City, Iowa 52243-0168

Table of Contents

CHAPTER 1 THE ACT® 1

ACT's Mission *1*
1.1 Philosophical Basis for the ACT *1*
1.2 Overview of the ACT *2*
1.3 Purposes, Claims, Interpretations, and Uses of the ACT *3*
1.4 Evidence-Based Design of the ACT Test *5*
1.5 ACT's Commitment to Fair Testing *7*
1.6 The Population Served by the ACT *8*
1.7 Test Preparation *9*

CHAPTER 2 THE ACT TEST DEVELOPMENT 10

2.1 Overview *10*
2.2 Description of the ACT Tests *10*
2.3 The ACT National Curriculum Survey *11*
2.4 Test Development Procedures *13*
2.5 Test Development Procedures for the Writing Test *21*
2.6 ACT Scores *22*

CHAPTER 3 CONTENT SPECIFICATIONS 25

3.1 Overview *25*
3.2 English Test *25*
3.3 Mathematics Test *27*
3.4 Reading Test *32*
3.5 Science Test *35*
3.6 Writing Test *38*

CHAPTER 4 TEST ADMINISTRATION, TEST SECURITY, AND ACCESSIBILITY AND ACCOMMODATIONS 41

4.1 Test Administration Overview *41*
4.2 Test Security *43*
4.3 Test Administration and Accessibility Levels of Support *47*

CHAPTER 5 SCORING AND REPORTING 50

5.1 Overview *50*
5.2 Test Section, Composite, STEM, and ELA Scores *50*
5.3 Detailed Performance Description *64*
5.4 Progress Toward the ACT WorkKeys National Career Readiness Certificate Indicator *68*
5.5 ACT College and Career Readiness Standards *69*
5.6 ACT College Readiness Benchmarks *76*

CHAPTER 6 SCALING, EQUATING, AND TECHNICAL CHARACTERISTICS 79

6.1 Scaling and Equating of the ACT English, Mathematics, Reading, and Science Tests *79*
6.2 Scaling and Equating of the ACT Writing Test for ACT ELA Score Calculation *83*
6.3 Reliability and Measurement Error *83*
6.4 Mode Comparability for Online Testing *94*

CHAPTER 7 VALIDITY EVIDENCE FOR THE ACT TESTS 99

7.1 Using ACT Scores to Measure Educational Achievement *99*

REFERENCES 103

APPENDIX A 107

List of Tables

TABLE 1.1. DEMOGRAPHIC CHARACTERISTICS OF WISCONSIN STATE CONTRACT SPRING 2025 ACT TESTERS 8
TABLE 2.1. ACT NATIONAL CURRICULUM SURVEY 2020 RESPONDENTS 13
TABLE 2.2. DIFFICULTY[A] DISTRIBUTIONS AND MEAN DISCRIMINATION[B] INDICES FOR ACT TEST ITEMS, JUNE 2023 TO APRIL 2024 18
TABLE 2.3. ACT TEST ITEMS EXHIBITING DIF BASED ON 2025 WISCONSIN STUDENT DATA 20
TABLE 3.1. DOK LEVEL DESCRIPTIONS FOR ENGLISH 26
TABLE 3.2. SPECIFICATION RANGES BY REPORTING CATEGORY FOR ENGLISH 27
TABLE 3.3. DOK LEVEL DESCRIPTIONS FOR MATHEMATICS 28
TABLE 3.4. SPECIFICATION RANGES BY REPORTING CATEGORY FOR MATHEMATICS 32
TABLE 3.5. DOK LEVEL DESCRIPTIONS FOR READING 33
TABLE 3.6. SPECIFICATION RANGES BY REPORTING CATEGORY FOR READING 34
TABLE 3.7. DOK LEVEL DESCRIPTIONS FOR SCIENCE 36
TABLE 3.8. SPECIFICATION RANGES BY REPORTING CATEGORY FOR SCIENCE 37
TABLE 3.9. SPECIFICATION RANGES BY SCIENCE CONTENT AREA 38
TABLE 3.10. DOK LEVEL DESCRIPTION FOR WRITING 38
TABLE 5.1. WRITING TEST ANALYTIC SCORING RUBRIC 54
TABLE 5.2. SAMPLE OF QUALITY REPORTS 58
TABLE 5.3. SUMMARY STATISTICS OF THE ACT TEST SCORE DISTRIBUTIONS FOR THE PRIMARY FORM ADMINISTERED IN WISCONSIN IN SPRING 2025 62
TABLE 5.4. SUMMARY STATISTICS OF THE ACT TEST SCORE DISTRIBUTIONS FOR THE ACCOMMODATED FORM ADMINISTERED IN WISCONSIN IN SPRING 2025 62
TABLE 5.5. SCALE SCORE COVARIANCES FOR MULTIPLE-CHOICE TESTS FROM THE PRIMARY FORM ADMINISTERED IN WISCONSIN IN SPRING 2025 64
TABLE 5.6. EFFECTIVE WEIGHTS OF THE ACT TESTS FROM THE PRIMARY FORM ADMINISTERED IN WISCONSIN IN SPRING 2025 64
TABLE 5.7. CORRELATIONS AMONG THE ACT TEST SCORES ADMINISTERED IN WISCONSIN IN SPRING 2025 64
TABLE 5.8. SUMMARY STATISTICS OF THE ACT WRITING AND WRITING DOMAIN SCORE DISTRIBUTIONS FOR THE WISCONSIN SPRING 2025 ADMINISTRATION 67
TABLE 5.9. CORRELATIONS AMONG THE ACT WRITING AND WRITING DOMAIN SCORES FOR THE WISCONSIN SPRING 2025 ADMINISTRATION 67
TABLE 5.10. ILLUSTRATIVE LISTING OF MATHEMATICS ITEM DIFFICULTIES BY SCORE RANGE 72
TABLE 5.11. NUMBER OF ACT ITEMS REVIEWED DURING 1997 NATIONAL REVIEW 72
TABLE 5.12. PERCENTAGE OF AGREEMENT OF 1997 NATIONAL EXPERT REVIEW 75
TABLE 5.13. ACT COLLEGE READINESS BENCHMARKS 77
TABLE 6.1. SCALE SCORE RELIABILITY AND SEM FOR THE WISCONSIN SPRING 2025 ACT TEST SCORES 85
TABLE 6.2. RAW SCORE RELIABILITY AND SEM FOR THE WISCONSIN SPRING 2025 ACT REPORTING CATEGORIES 85
TABLE 6.3. AGREEMENT RATES FOR THE ACT WRITING DOMAIN SCORES FOR THE WISCONSIN SPRING 2025 ADMINISTRATION 89
TABLE 6.4. CLASSIFICATION CONSISTENCY FOR THE WISCONSIN SPRING 2025 PERFORMANCE LEVEL CUT SCORES 92

TABLE 6.5. CLASSIFICATION CONSISTENCY FOR THE WISCONSIN SPRING 2025 ACT READINESS RANGES 93
TABLE 6.6. COMPOSITE SCORE RANGES FOR THE ACT NCRC LEVELS 94

LIST OF FIGURES

FIGURE 1.1. THE FULL PICTURE: EVIDENCE AND VALIDITY 7
FIGURE 5.1. OVERALL SCORE AND PERCENTILE RANK ON A SAMPLE INTERACTIVE SCORE REPORT ON MYACT 51
FIGURE 5.2. OVERALL SCORE AND PERCENTILE RANK ON A SAMPLE ACT HIGH SCHOOL SCORE REPORT 52
FIGURE 5.3. DETAILED RESULTS ON A SAMPLE INTERACTIVE SCORE REPORT ON MYACT 65
FIGURE 5.4. DETAILED RESULTS ON A SAMPLE ACT HIGH SCHOOL SCORE REPORT 66
FIGURE 6.1. CSEM FOR MULTIPLE-CHOICE TEST SCORES 86
FIGURE 6.2. AVERAGE AND FITTED CSEMS FOR ACT WRITING TEST SCALE SCORES 88
FIGURE 6.3. CSEM FOR COMPOSITE SCORES 90
FIGURE 6.4. CSEM FOR STEM SCORES 91
FIGURE 6.5. CSEM FOR ELA SCORES 91

Chapter 1
The ACT®

ACT's Mission

ACT has been dedicated to improving college and career readiness for all students since its inception in 1959. ACT's renowned longitudinal system of assessments, with the ACT® test as a capstone, has provided students, educators, and policymakers with unparalleled measures of college and career readiness. ACT's mission is helping people achieve education and workplace success.

1.1 Philosophical Basis for the ACT

Underlying the ACT is the belief that students' preparation for college and the workplace is best assessed by measuring, as directly as possible, the skills learned in high school that are required for success in college-level courses. The required academic skills can be assessed most directly by reproducing, as faithfully as possible, the complexity of the work students do in the classroom. Therefore, ACT's tests of educational achievement are designed to determine how skillfully students solve problems, grasp implied meanings, draw inferences, evaluate ideas, and make judgments in subject-matter areas important to success in college.

The ACT is oriented toward the general content areas of college and high school instructional programs. The test questions require students to integrate the knowledge and skills they possess in major curriculum areas with the information provided by the test. Thus, scores on the test are directly related to the students' educational progress in curriculum-related areas and possess meaning that is readily grasped by students, parents, and educators.

The constructs measured by the ACT section tests are supported by multiple sources of validity evidence (see *Chapter 7*). For example, ACT has, for many years, collected longitudinal statistical evidence backing the strong relationship between student performance on the section tests and student performance in entry-level courses in the corresponding subjects. More recent methodologies such as cognitive labs have served to further confirm this evidence.

Because tests of educational achievement measure many of the skills taught in high school, the best preparation for achievement tests is rigorous high school coursework. Long-term learning in school, rather than short-term cramming and coaching, becomes the obvious best form of test preparation. Thus, educational achievement tests serve as motivators by sending students a clear message that high test scores reflect not simply innate ability but a level of achievement that has been reached as a result of hard work.

The ACT requires students to apply critical thinking skills when comprehending complex texts, analyzing data displays showing the results of scientific experiments, producing effective argumentative writing, and solving sophisticated mathematics problems. Therefore, in order to acquire such skills and achieve high scores on the ACT, students may be influenced to choose

challenging coursework in high school. In this way, the ACT may help high schools develop their students' critical thinking skills, which will be important for success in college and later life. Thus, the ACT is designed not only to accurately reflect educational goals that are widely considered important by educators, but also to emphasize the importance of a student's educational decisions.

1.2 Overview of the ACT

The ACT emphasizes students' academic preparedness by directly addressing the content domains students must master to achieve college and career readiness. The main component of the ACT is a standardized battery of four tests of educational achievement—English, mathematics, reading, and science—along with an optional writing test. Through ACT's online registration and data collection system (MyACT), ACT also collects information about students' high school courses and grades, educational and career aspirations, extracurricular activities, and educational needs.

The ACT provides information about how well a student performs compared to other students. It also provides standards-based interpretations through ACT's *College and Career Readiness Standards* (CCRS)—empirically derived descriptions of the essential skills and knowledge students need in order to become ready for college and career success. Using the CCRS, secondary educators can pinpoint the skills students have and those they are ready to learn next. The CCRS clarify college expectations in terms that high school teachers understand. The CCRS also offer teachers guidance for improving instruction to help correct student deficiencies in specific areas. ACT's *College Readiness Benchmarks* are the minimum scores associated with a high likelihood of postsecondary success in each content area. Together, the CCRS and the Benchmarks provide students specific insights to support success in college and career. *Chapter 5* gives details about the CCRS and Benchmarks.

1.3 Purposes, Claims, Interpretations, and Uses of the ACT

The purposes, claims, interpretations, and uses of the ACT are reflected in a theory of action that integrates evidence supporting content validity (academic research, curriculum information, and academic standards) with predictive validity (empirical data). The theory of action begins by answering fundamental questions about the purpose, users, uses, benefits, claims, interpretations, and outcomes of the test.

Intended Purpose. The primary purpose of the ACT is to measure students' level of college and career readiness in core academic areas. ACT testing is intended to help high school students develop postsecondary educational plans and to help postsecondary educational institutions meet the needs of their students.

In service of the intended purpose, the ACT provides an overall Composite score and scores for each of the section tests and the optional writing test. The test also provides a measure of students' STEM (science, technology, engineering, and math) skills (by combining mathematics and science scores), an Understanding of Complex Texts (UCT) indicator, and an ELA (English language arts) score (by combining English, reading, and writing scores; only students who take the writing test can receive an ELA score). The test also provides information about student achievement at a more detailed level through the reporting category scores on each test section.

Intended Users. Primary intended users of the ACT test include high school students (typically in Grades 11 and 12), the educational agencies or organizations supporting the academic preparation of these students (i.e., schools, districts, and states), postsecondary institutions, and talent recognition and scholarship agencies.

Intended Uses. ACT test data, test scores, and score interpretations have several intended uses. Students use their results to plan for further education and explore careers based on their skills, interests, and aspirations. High schools use ACT data in academic advising and counseling, evaluation studies, accreditation documentation, and public relations. Postsecondary institutions use ACT results to support admission and course placement decisions. States use the ACT as part of their statewide assessment systems to measure students' educational achievement, to monitor educational improvement and achievement gaps over time, and to meet federal accountability requirements. Many private, state, and national agencies that provide scholarships, loans, and other types of financial assistance use ACT test scores to help assess students' academic qualifications. Agencies also use ACT data to identify academically talented students as early as middle school.

Intended Benefits. The ACT test benefits its users by:

- allowing students to demonstrate the knowledge and skills gained throughout educational coursework in English, mathematics, reading, science, and writing;

- providing students with a profile of their relative strengths and weaknesses in the subject areas assessed by the test, thereby informing students about what they know and can do (based on the College and Career Readiness Standards);
- providing parents with insights about their students' knowledge and skills;
- providing educators (in schools, districts, and states) with information about their students' knowledge and skills;
- encouraging students to better prepare for college and careers by taking courses linked to positive postsecondary outcomes;
- indicating whether a student is likely ready for college-level coursework or a work training program (based on the College and Career Readiness Benchmarks and the Progress Toward the ACT® WorkKeys® National Career Readiness Certificate® (NCRC®) indicator); and
- providing colleges and talent identification and scholarship agencies with information about students' level of achievement in the subject areas assessed by the test.

Interpretations and Claims. The interpretations and claims of the ACT include the following:

- The ACT measures academic knowledge and skills that are acquired in high school and are important for college-level coursework in English, mathematics, reading, science, and writing.
- ACT scores can be used in combination with other relevant measures to estimate students' likelihood of success in college during the first year and beyond and to help inform college admission, course placement, and remediation decisions.
- ACT scores can be used in aggregate for monitoring educational improvement and achievement gaps over time, as well as assisting with evaluating the effectiveness of school and district programs when a school administers the ACT to all its students.
- MyACT includes the ACT Interest Inventory (ACT, 2009b), which is based on research about career planning, to point students toward a range of good-fit options to consider. In the process of exploration, students can focus on educational and occupational options that are relevant to future satisfaction and success. The ACT Interest Inventory results, when used in conjunction with ACT test scores, provide a more holistic picture of the student's educational development and career-relevant motivations.

Intended Outcomes. Using the results of the ACT in conjunction with other academic and non-academic measures can help

- students, parents, and educators to identify academic knowledge and skills in which students might benefit from additional instruction and supports while still in high school to better prepare for college and career and avoid taking remedial or developmental courses in their first year of college;
- students to expand their educational and occupational exploration beyond options initially considered based on students' academic strengths and weaknesses and interests measured by the ACT Interest Inventory (ACT, 2009b) or through ACT's Educational Opportunity Service (Moore & Cruce, 2017);
- schools and districts to raise college awareness and exposure when all students take the ACT through state or district testing;
- schools and districts to evaluate student growth and identify gaps in educational achievement in order to better understand which school programs are effective in preparing all students for college and career;
- postsecondary institutions to select students for admission who are likely to enroll at the institution and, once enrolled, likely to succeed in their college courses and complete a college degree at the institution;
- postsecondary institutions to place students in first-year college courses in which they are most likely to be successful; and
- postsecondary institutions to identify students early on who are most likely to struggle academically, who may be at risk of dropping out of college, and who may benefit from institutional academic services and supports in order to successfully transition from high school to college.

1.4 Evidence-Based Design of the ACT Test

The design of the ACT test emerges from an evidence-based research and data collection process that ensures that items and test forms elicit the evidence necessary to support the claims of the ACT. For example, content and item specifications and test blueprints influence the technical quality of test items and forms. The ACT design is informed by several factors, including the following:

- Subject-matter experts (SMEs)
- Academic research on skill targets, sequencing of skills, and grade placement
- Data and evidence of student understanding collected from the ACT test
- The ACT® National Curriculum Survey®

- A survey of standards frameworks—including, but not limited to, the ACT College and Career Readiness Standards, the Next Generation Science Standards, and other college and career readiness standards

The validity argument is further supported with criterion-related longitudinal evidence from students who complete the ACT and then go on to colleges (two-year and four-year) and career-training programs.

While SMEs can identify copious skills covered by a typical high school curriculum, not all skills and knowledge are essential for postsecondary success, nor will measuring every skill help identify lower- and higher-achieving students. For example, some skills essential for success may be attained by more than 95 percent of students continuing on to postsecondary education, and including items that measure such skills on a test only increases test length without contributing to predicting postsecondary success.

Similarly, ACT research demonstrates that there are often discrepancies between skills high school educators see as relevant to success and the expectations and experience of college faculty. Again, ACT uses data from a national sample of institutions, academic programs, and college majors to prioritize the skills and knowledge clearly linked to student success.

ACT supplements these other sources of data with subject-matter expertise. ACT's test development staff has extensive classroom experience in the subjects tested by the ACT.

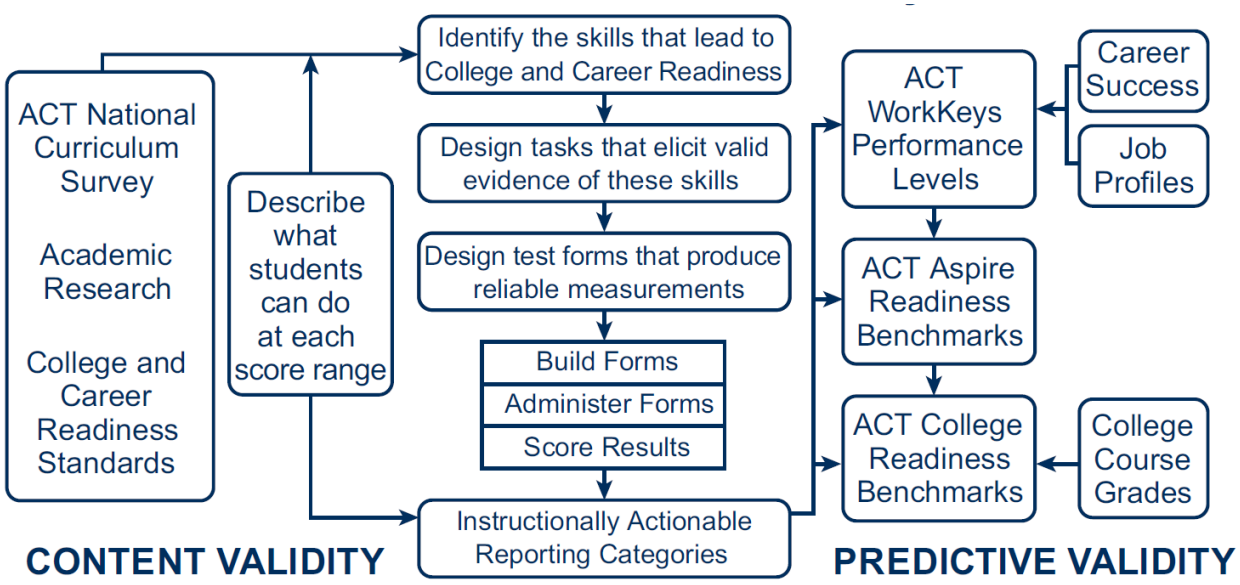
The first step in developing the ACT was to synthesize research on high-value skill targets—the skill targets that can be shown to offer the most useful evidence of college and career readiness. This evidence was obtained by organizing the knowledge and skills identified by educators and contained in educational standards into the assessment content framework.

The next step was to use this research to develop content specifications and task models that articulated the evidence needed to monitor student progress. Tasks were then generated from these specifications and assembled into test forms based on test blueprints.

The test blueprints specify constraints on various factors, including, but not limited to, content coverage, item difficulty, cognitive complexity, reading load, and the time required for an item. Test forms are then administered, and student performance data are collected.

Figure 1.1 helps illustrate how a validity argument is composed of multiple sources of research, empirical data, and other forms of evidence. Content validity is shown to be based in research. Predictive validity information flows in primarily from the ACT and, to a lesser extent, the ACT® WorkKeys® assessments. Both channels supply information about which knowledge and skills are needed to perform well on the ACT, thus supporting an iterative model of refinement that serves the common goal of determining whether a student is college and career ready.

Figure 1.1. The Full Picture: Evidence and Validity



1.5 ACT's Commitment to Fair Testing

Fairness is an essential quality of testing related to issues such as testing experience, possible measurement bias, equitable score interpretations, and students' ability to accurately demonstrate the extent of their knowledge and skills (i.e., accessibility). Since publication of the original edition in 1988, ACT has endorsed the *Code of Fair Testing Practices in Education* (*Code*; Joint Committee on Testing Practices, 2004), a statement of the obligations to test takers of those who develop, administer, or use educational tests and test data. The development of the *Code* was sponsored by a joint committee including the American Counseling Association, the American Educational Research Association, the American Psychological Association, the American Speech-Language-Hearing Association, the National Association of School Psychologists, the National Association of Test Directors, and the National Council on Measurement in Education, to advance, in the public interest, the quality of testing practices.

The *Code* sets forth fairness criteria in four areas: developing and selecting appropriate tests, administering and scoring tests, reporting and interpreting test results, and informing test takers. Separate standards are provided for test developers and for test users in each of these four areas. According to the *Code*, for example, test developers should provide "tests that are fair to all test takers regardless of age, gender, disability, race, ethnicity, national origin, religion, sexual orientation, linguistic background, or other personal characteristics" (Code, 2004, p. 2). Test developers should "avoid potentially offensive content or language" (Code, 2004, p. 4) and "evaluate the evidence to ensure that differences in performance are related to the skills being assessed" (Code, 2004, p. 4). ACT's endorsement of the *Code* represents a commitment to vigorously safeguarding the rights of individuals participating in its testing programs.

Similarly, ACT endorses, and is committed to complying with, *the Code of Professional Responsibilities in Educational Measurement* (NCME Ad Hoc Committee on the Development of a Code of Ethics, 1995), a statement of professional responsibilities for those who develop assessments; market and sell assessments; select assessments; administer assessments; interpret, use, and communicate assessment results; educate about assessments; and evaluate programs and conduct research on assessments. One of those responsibilities is to "develop assessment products and services that are as free as possible from bias due to characteristics irrelevant to the construct being measured" (*Section 1.2*).

Ensuring fairness in a test is a critically important goal. Lack of fairness must be detected, eliminated, and prevented at all stages of test development, test administration, and scoring. The work of ensuring fairness starts with the design of the test and test specifications. It then continues through every stage of the test development process, including item writing and review, item pretesting, item selection and forms construction, and forms review. ACT makes every effort to see that ACT tests are fair to the populations for which the tests are intended and is committed to participating in ongoing dialogues about assessment fairness.

1.6 The Population Served by the ACT

During the Spring of 2025, WDPI contracted with ACT to provide the ACT with writing to public school and parental choice 11th-grade students during regular school hours at schools certified as ACT state testing sites. The analyses reported in this technical manual are based on 63,361 student records from the spring 2025 administration of the ACT in the state-sponsored school-day testing in Wisconsin.

Table 1.1. Demographic Characteristics of Wisconsin State Contract Spring 2025 ACT Testers

Demographic	Percentage	N
Gender		
Female	47	29,688
Male	49	31,143
Other Gender	1	388
No response	1	905
Prefer not to respond	2	1,237
Racial/Ethnic Background		
African American/Black	7	4,379
White	64	40,262
American Indian/Alaska Native	1	525
Hispanic/Latino	14	8,702
Asian	4	2,486
Native Hawaiian/Other Pacific Islander	< 1	49
Two or more races	5	3,164
Prefer no response/blank	6	3,794

Notes: Due to rounding, some values may not add to exactly 100%. Information in this table can also be found in *Appendix Table 1*.

1.7 Test Preparation

Awareness of and exposure to an assessment prior to taking it is important in order for students to feel comfortable and confident. ACT offers a variety of free and affordable test preparation solutions for students, parents, and educators.

- ***Preparing for the ACT Test.*** Includes a full-length practice test, test-taking strategies, and information about what to expect on test day. This publication is available in English and Spanish as a free download for teachers, students, parents, and others.
  - English: *www.act.org/content/dam/act/unsecured/documents/Preparing-for-the-ACT.pdf*
  - Spanish: *https://www.act.org/content/dam/act/unsecured/documents/Preparing-for-the-ACT-Spanish.pdf*
- ***ACT Official Online Practice Test.*** ACT provides free access to a full-length practice test that simulates an online testing experience. Students may access both timed and untimed practice tests for each test section. Students may sign into each of the section tests as often as they wish in order to become comfortable with the testing.
- ***Alternate Assessment Format Samples.*** Students who will test with alternate formats of the assessment can prepare by practicing with one of our alternate format samples. Braille, large print, audio, and reader's script formats are available at no cost and contain a full-length practice test.

Chapter 2
The ACT Test Development

2.1 Overview

This chapter describes ACT's test development process—including item and form development procedures. The following principles have shaped and will continue to drive ACT's development agenda:

1. Report results in instructionally relevant ways that support clear interpretation within content areas.

2. Maintain reasonable testing times by assessing what research and evidence show to be the most critical factors for success after high school.

3. Leverage technology to enhance student engagement, produce more meaningful results, and share results in a timely fashion.

4. Increase the emphasis on evidence-centered design, implement best practices as they mature, and improve ACT's capabilities to enact the highest-quality design and development processes.

5. Include science as a core academic domain in ACT's assessment batteries.

6. Reflect the research-validated reality that there are multiple dimensions of readiness and success.

As an educational research organization, ACT uses these principles to drive the development and continuous improvement of ACT's education and workplace solutions, as well as the research agenda associated with them, thereby enabling ACT to fulfill its mission of helping all individuals achieve education and workplace success.

This chapter provides brief overviews of the ACT® National Curriculum Survey®, the content and bias review process, and the statistical criteria for selecting operational items and assembling forms. This chapter concludes with a high-level explanation of the ACT scoring procedures, including descriptions of additional scores and indicators.

2.2 Description of the ACT Tests

The ACT® test contains four sections—English, mathematics, reading, and science—and an optional writing test. These tests measure important content, skills, and concepts taught in high school and needed for success in college and career. The content specifications describing the knowledge and skills to be measured by the ACT were determined through a detailed analysis of relevant information. ACT uses direct feedback from current high school and postsecondary teachers (via the ACT National Curriculum Survey, as well as through external review of test

items) and student data from the ACT and from grades earned in postsecondary courses. These data are used to verify that the ACT measures knowledge and skills empirically linked to postsecondary and career success. The ACT National Curriculum Survey is described in the subsequent section of this chapter. Information about the specific knowledge and skills measured by each test is provided in *Chapter 3*. *Chapter 7* describes sources of validity evidence supporting the interpretation of ACT scores.

2.3 The ACT National Curriculum Survey

The ACT National Curriculum Survey is a one-of-a-kind nationwide survey, conducted by ACT every few years, of educational practices and college and career readiness expectations (ACT, 2007, 2009a, 2013a, 2016a, 2020). The ACT National Curriculum Survey embodies ACT's commitment to ensuring not only that the assessments are consistently valid and relevant but also that they provide information enabling students and workers to be fully ready to embark successfully on rewarding college and career journeys.

ACT surveys thousands of K–12 teachers and college instructors in English and writing, mathematics, reading, and science, as well as a national cross section of workforce supervisors and employees, for the purpose of determining which skills and knowledge in these subjects are currently being taught at each grade level and which skills and knowledge are currently considered essential aspects of college and career readiness.

Questions are also included about which skills from the ACT® Holistic Framework®—a research-based framework that integrates behavioral skills, education and career navigation skills, core academic skills, and cross-cutting capabilities (such as teamwork and critical thinking)—are most integral to college and career success.

ACT uses the results of the ACT National Curriculum Survey to guide the development of ACT assessment solutions, including the ACT test, the PreACT®, the PreACT® Secure, and ACT® WorkKeys®. ACT conducts the survey to ensure that its assessments are measuring the knowledge and skills that instructors of credit-bearing, first-year college courses identify as important for success in each content area or that workforce supervisors identify as important for readiness for targeted workforce training and for success on the job.

ACT makes the results of each ACT National Curriculum Survey public to help education and workforce stakeholders make more informed decisions about the skills needed to be successful in postsecondary education and the workplace.

2.3.1 The Purpose of the ACT National Curriculum Survey

The ACT National Curriculum Survey is a crucial step in the process of building and regularly updating a suite of ACT assessments that is empirically aligned to college readiness standards. Survey results help address a critical question: Does the test measure knowledge and skills currently relevant to college and career success? Ultimately, the survey data inform the blueprints for the assessments. Subsequently, results from the assessments are used to

validate ACT's College and Career Readiness Standards as well as its College and Career Readiness Benchmarks.

Equally important is predictive validity. Using postsecondary course performance data, ACT answers a second critical question: Does the test accurately predict postsecondary performance? Constant monitoring allows ACT to ensure that the answer to both questions is "yes."

ACT uses the findings from the ACT National Curriculum Survey to monitor the test blueprints. This process ensures that the assessments measure not only what is being taught in schools around the country but also what demonstrably matters most for college and career readiness. To maintain relevancy and currency, it is important that assessments be built upon up-to-date evidence of what knowledge and skills matter most according to the assessment context and purpose.

The science behind ACT assessments—that is, the evidence base and ongoing research—is critical to answering the key question of what matters most for college and career readiness. The ACT National Curriculum Survey represents ACT's commitment to

- use evidence and research to develop and validate ACT standards, assessments, and benchmarks;
- maintain a robust research agenda to report on key educational metrics; and
- develop assessments, reports, and interventions that will help individuals navigate their personal path to success along the kindergarten-through-career continuum.

2.3.2 Survey Sample and Process

For the 2020 ACT National Curriculum Survey, ACT recruited participants via various print and electronic methods (e.g., advertisements, email, social media) and invited participation from educators at the early elementary school, late elementary school, middle school, high school, and college levels who teach courses in English and writing, mathematics, reading (including English language arts and social studies), and science (including biology, chemistry, physics, and earth and space science) in public and private institutions across the United States. ACT also invited participation from supervisors and employees at a large variety of businesses. Table 2.1 gives the number of survey respondents in each area.

Table 2.1. ACT National Curriculum Survey 2020 Respondents

Area	Number of Respondents
Early elementary school	1,214
Late elementary school	1,213
Middle school	1,623
High school	1,619
K–12 administrators	405
College instructors	2,883
Workforce supervisors	405
Workforce employees	406
Total	9,768

Education participants were asked to rate discrete content knowledge and skills with respect to how important each is to student success in the content area. Specifically, K–12 teachers were asked to rate the importance of content knowledge and skills in a given class they teach, while college instructors were asked to rate the importance of content knowledge and skills as prerequisites to success in a given class they teach.

ACT also asked the K–12 teachers to indicate whether they teach particular content knowledge or skills and, if so, whether those knowledge or skills are taught as standard parts of their courses or as part of a review of materials that should have been learned earlier. Some education participants were also asked other content-related questions depending on the grade level they taught.

Workforce participants were asked to rate discrete skills with respect to how important each is to success in entry-level positions. ACT also asked workforce participants to indicate how often employees in their workplace use each of these skills on the job.

Finally, ACT asked all participants questions relevant to current education policy issues (e.g., assessments, technology, standards, student characteristics, and obstacles to success). All results are discussed in the report for the ACT National Curriculum Survey 2020 (ACT, 2020). To ensure that no single content area would have more influence than another on results, the educational-level totals were averaged across English language arts, mathematics, and science.

2.4 Test Development Procedures

2.4.1 Test Specifications Overview

As described below, two major types of test specifications are used in developing the ACT tests: content specifications and statistical specifications. Several other considerations are made when new test forms are created, such as meeting passage and item word count requirements, avoiding very long strings of the same response option, and preventing extreme imbalance in the distribution of response options.

Content specifications. Content specifications for the ACT tests were developed through the curricular analysis discussed above. Those specifications define the approximate number of items from each reporting category and cognitive complexity level on a test form. They also set expectations for diverse representation in passages in terms of gender, ethnicity, region, and community type (urban or rural). To support validity and fairness, ACT ensures that the content specifications include only knowledge and skills aligned to the intended purposes of the test. To include anything else in the content specifications would invite construct-irrelevant variance that could unfairly impact students' scores. While care is taken to ensure that the basic structure of each ACT test remains the same from year to year, the specific characteristics of the test items used in each specification category are reviewed regularly. While the general content of the test remains constant, the particular kinds of items in a specification category may change slightly. The basic content structure of each ACT test is provided in *Chapter 3*.

Statistical specifications. Statistical specifications for the tests indicate the average level of item difficulty (proportion correct), the distribution of item difficulties, and the minimum acceptable level of discrimination (biserial correlation) of the test items to be used.

The tests are constructed with a certain target mean item difficulty for the ACT population in each subject area. Individual item difficulty must fall within a range from about 0.15 to 0.89 for mathematics and about 0.20 to 0.85 for English, reading, and science. The difference mainly reflects the fact that mathematics items have five answer options, but other items offer only four answer options. The statistical specifications also prescribe approximate numbers of items with difficulties falling in certain ranges (0.10–0.19, 0.20–0.29, and so forth), which ensures that each test form includes a mix of low-, moderate-, and high-difficulty items. This specification helps ensure that test scores are reliable for students across the spectrum of achievement levels.

With respect to discrimination indices, items should have a biserial correlation of 0.20 or higher with test scores measuring comparable content. Thus, for example, performance on mathematics items should correlate 0.20 or higher with overall performance on the mathematics test. Such items help identify students with lower and higher levels of achievement, thereby contributing to the reliability of test scores.

2.4.2 Item Writers

ACT relies primarily on internal content specialists to develop items. Content specialists are subject matter experts, trained in the disciplines for which they write items. Most have experience in teaching at various levels, from high school to university, and at a variety of institutions, from small private schools to large public institutions. ACT makes every attempt to include item writers who represent the diversity of the population of the United States with respect to ethnic background, gender, and geographic location.

Each content specialist is familiar with an item writer's guide that is specific to the content area. The guides include example items, test specifications, and ACT's requirements for content and style. Also included are specifications for the fair portrayal of all groups, which includes avoidance of subject matter that may be unfamiliar to members of certain groups within society,

a balanced representation of race/ethnicity, and gender-neutral language. Item development assignments are balanced among content specialists to ensure a diversity of material.

Depending on development needs, ACT may contract with external item writers or make use of automated item generation. Externally contracted item writers are also specialists in the content areas measured by the test and typically have teaching experience. Each potential item writer is required to submit a sample set of materials (written using the item writer's guide) for ACT's evaluation. Item writers contracted with ACT are held to the same high-quality standards as internal content specialists, and the same attempts to maintain diversity of material and security of the testing program are made. Automated item generation makes use of models with interchangeable elements based on items that were administered in the past and exhibited desirable statistical properties (e.g., difficulty and discrimination).

2.4.3 Item Writing

Item-writing assignments are driven by the test blueprint and item pool analyses, with the goal of attaining a wide range of high-quality items to elicit evidence of the knowledge, skills, and abilities measured in each test. A typical assignment is tied to an evidenced-based item template and focuses on a skill statement that the item needs to assess. Included in each template is a set of statements describing what evidence of students' knowledge and skills should be elicited by the item.

Assignments are constructed through ACT's item authoring system. This system also contains item metadata, information about the item flow through the stages of development, comments from reviewers, and item quality metrics.

All items must be educationally important and psychometrically sound. Many items must be constructed because, even with good writers, many pretested items fail to meet ACT's standards.

Each item writer submits a set of items in a given content area. All mathematics items developed recently are discrete (not passage based); some older items belong to a set (i.e., several items based on the same paragraph or chart). All items on the English and reading tests are related to prose passages. Some reading items may be related to visual or quantitative information, such as graphs and tables, attached to a passage. All items on the science test are related to passages that contain data presentations such as graphs and tables.

2.4.4 Review of Items

Content Review

After an item (or set of items) is written, it is reviewed several times by numerous content specialists to verify that it meets all of ACT's standards. It is edited to meet ACT's specifications for content accuracy, word count, item classification, item format, and language. During the review and editing process, all test materials are reviewed for fair portrayal and balanced representation of groups within society and for gender-neutral language.

After internal item reviews are completed, ACT invites external reviewers with knowledge and experience in those content areas, including practicing secondary and postsecondary educators, to participate in refining items and verifying that they should elicit evidence of the intended constructs. During external review, every item is independently reviewed by four to six subject matter experts from across the United States, each of whom has extensive experience with students at or around the grade levels at which the test content is typically taught. During the external content review, items are evaluated for content accuracy, item format, and the effectiveness of language in terms of leveling, precision, and fairness.

Fairness Reviews

Fairness reviews play an essential role in the development of ACT assessments. In order to help ensure that content is fair, unbiased, and accessible, we conduct external fairness reviews for all items prior to pretesting and for entire test forms before they become operational. In this context, “accessible” means that examinees can access the construct measured by the assessment and accurately demonstrate their construct-relevant knowledge and skills when responding to test items. Avoiding content that is potentially biased is one important aspect of accessibility. *Chapter 4* describes ACT’s approach to another aspect of accessibility: designing tests and providing testing accommodations for English learners and students with disabilities.

The external fairness review panel consists of experts in diverse areas of education who have experience working with diverse populations. Passages and items are reviewed to help verify that content is not unfair, biased, or insensitive. All comments are reviewed by ACT content specialists, and appropriate changes are made. For both content reviews and fairness reviews, we select reviewers so that no one state is overrepresented, because our stakeholders count on national representation to maintain the comparability of test scores.

2.4.5 Item Tryouts

ACT pretests every item before it appears on an operational form to verify that the item functions properly—that is, the item is not too easy or difficult, the item contributes to precise measurement of the intended construct, and there are no problems with the correct response or distractors. Items and passages that are judged to be acceptable in the review process are assembled into tryout units (compilations of items and any associated passages). These tryout units are then appended to paper test booklets administered during Saturday national testing events. Each examinee is administered a tryout unit from one of the four academic areas covered by the ACT tests, with the exception of the writing test, which is pretested in a separate standalone tryout. The tryout unit is sometimes referred to as the fifth test in the ACT battery, though performance on the tryout items does not affect examinees’ ACT scores. The tryout units are spiraled so that each unit is administered to a random sample of examinees participating in a given administration, which helps ensure that the psychometric properties of the items—especially item difficulty—are comparable across items and that all item statistics reflect performance from representative samples of examinees.

Item Analysis of Tryout Units

Item analyses are performed on the tryout units. For a given unit, the sample is divided into low-, medium-, and high-performing groups by the individuals' scores on the ACT test in the same content area (taken at the same time as the tryout unit). The cutoff scores for the three groups are the 27th and the 73rd percentiles in the distribution of those scores. These percentiles maximize the critical ratio of the difference between the mean scores of the upper and lower groups, assuming that the standard error of measurement in each group is the same and that the scores for the entire examinee population are normally distributed (Millman & Greene, 1989).

Proportions of students in each of the groups correctly answering each tryout item are tabulated, as are the proportions in each group who select each of the incorrect options. The biserial and point-biserial correlation coefficients of each tryout item are also computed.

Item analyses identify statistically effective test items. Items that are either too difficult or too easy are eliminated or revised for future item tryouts, as are items that fail to discriminate between students of high and low educational achievement (as measured by their corresponding ACT test scores). The biserial and point-biserial correlation coefficients, as well as the differences between proportions of students answering the item correctly in each of the three groups, are used as indices of the discriminating power of the tryout items.

Additionally, differential item functioning (DIF) analysis is conducted on the tryout data. DIF can be described as a statistically significant difference between the odds of a certain group (the focal group) answering the item correctly and the odds of a comparison group (the reference group) answering the item correctly when students in the two groups have similar levels of achievement with respect to the content being tested. Items exhibiting DIF that is large in magnitude and statistically significant are examined by a diverse panel of external fairness reviewers, who evaluate whether there is a content-based explanation for the DIF.

Each item is reviewed following the item analysis. ACT staff members scrutinize items flagged for statistical reasons or DIF to identify possible problems. In some cases, items may be revised and sent through the tryout process again. The review process also provides feedback that helps to improve the quality of future items.

2.4.6 Assembly of New Forms

Items that are judged acceptable in the review process following item tryouts are placed in an item pool. Preliminary forms of the ACT tests are constructed by selecting from this pool items that match the content and statistical specifications (described in *Chapter 3*).

Table 2.2 displays the distributions of item difficulty levels on 7 forms administered from September 2024 to July 2025. In addition, mean point-biserial correlations and completion rates are reported. Table 2.2 indicates that the ACT forms included a small number of items with *p*-values falling outside the desired range of 0.15–0.89 for mathematics and 0.20–0.85 for English,

reading, and science. Such items were slightly easier or slightly more difficult than expected based on data from the item tryout stage.

The completion rate is an indication of whether a test is speeded for a group of students. A test is considered speeded if many students do not have sufficient time to answer the items in the time allotted. The completion rate reported in Table 2.2 for each test is the average completion rate for 7 national test dates from September 2024 to July 2025. The completion rate for each test is computed as the average percentage of examinees who answered all of the last five items.

Table 2.2. Difficulty[a] Distributions and Mean Discrimination[b] Indices for ACT Test Items, September 2024 to July 2025

Statistic	English	Mathematics	Reading	Science
Difficulty range				
.00–.09	0.00%	0.00%	0.00%	0.00%
.10–.19	0.00%	3.10%	0.36%	0.71%
.20–.29	2.48%	10.48%	1.79%	3.21%
.30–.39	3.81%	13.10%	3.93%	11.07%
.40–.49	9.14%	16.90%	15.00%	15.36%
.50–.59	18.10%	17.14%	22.14%	16.79%
.60–.69	27.81%	17.86%	25.71%	24.64%
.70–.79	23.43%	14.29%	18.93%	19.64%
.80–.89	14.29%	6.90%	11.07%	8.57%
.90–1.00	0.95%	0.24%	1.07%	0.00%
Number of scored items[c]	525	420	280	280
Mean difficulty	0.65	0.53	0.63	0.59
Mean discrimination	0.43	0.42	0.41	0.43
Mean completion rate[d]	95%	95%	96%	97%

[a] Item difficulty is the proportion of examinees who correctly answered the item.
[b] Item discrimination is the point-biserial correlation coefficient, which is also known as the item-total correlation.
[c] Each test form consists of 75 items for English, 60 for mathematics, 40 for reading, and 40 for science.
[d] Completion rate is the percentage of examinees who answered all of the last five items (averaged across forms).

2.4.7 Content and Fairness Review of Test Forms

The preliminary versions of the test forms are subjected to several reviews to ensure item quality and that the overall test forms meet content and statistical specifications and exemplify best practices supporting fair and accessible testing. ACT staff performs the first review. Items are checked for content accuracy and conformity to ACT style. The items are also reviewed to ensure that they are free of clues that could allow test-wise students to answer the items correctly even though they lack the required subject-area knowledge or skills. All ACT test forms

go through an external content review. Each form is reviewed by four to six educators from around the United States, each of whom has extensive experience with students at or around the grade levels at which the test content is typically taught. These reviews follow a process similar to the item development external content review. In addition to focusing on individual items, however, the reviewers also consider the quality of the form as a whole. They judge the form's distributions of content and cognitive complexity to make sure that there is no over- or under-representation in any category. Reviewers also look for the presence of cluing between items and other issues that could lessen the usefulness of the resulting scores.

Additionally, all newly developed ACT forms must go through external fairness reviews to support fair, equitable, and inclusive assessments that are accessible to all regardless of differences in background or perspective. As with the earlier fairness review, reviewers are experts in diverse areas of education who have experience working with diverse populations. At this stage, reviewers examine individual items and passages, but they also consider the preliminary form as a whole. That form should be balanced in multicultural and gender representation. While it is impossible, given the limited amount of material in each test form, to represent every group in every form, a good-faith effort to represent diversity should be discernable.

After the external reviews are complete, ACT summarizes the results. All comments from the consultants are reviewed by ACT content specialists, and appropriate changes are made to the test forms. Whenever significant changes are made, items and/or passages are replaced and are again reviewed by the appropriate consultants and by ACT staff. If no further changes are needed, the test forms are prepared for publishing.

2.4.8 Review Following Operational Administration

After each operational administration, item analysis results are reviewed for any anomalies, such as substantial changes in item difficulty and discrimination indices between tryout and operational administrations. Only after all anomalies have been thoroughly checked and the final scoring key approved are score reports produced. Examinees may challenge any items they feel are questionable. Once a challenge to an item is raised and reported, the item is reviewed by content specialists in the content area assessed by the item. In the event that a problem is found with an item, actions are taken to eliminate the influence of the problem item as necessary and appropriate. In all cases, each person who challenges an item is sent a letter indicating the results of the review.

Also, after each operational administration, differential item functioning (DIF) analysis is conducted on the test data. The procedure currently used for the analysis is the Mantel-Haenszel common odds ratio procedure (MH), which is also used during the pretest item analysis. The examinees' scores on each item are analyzed using the procedure to identify evidence of potential item bias. Items with MH statistics exceeding certain tolerance levels—determined based on pre-established criteria—are flagged. The flagged items can then be reviewed by content specialists for possible explanations of the MH results. In the event that a

problem is found with an item, actions can be taken to eliminate the influence of the problem item.

Table 2.3 lists the number of ACT items that exhibited DIF according to the MH procedure for forms administered during the 2024–2025 academic year based on Wisconsin students taking the state-sponsored ACT administration. Analyses were conducted to compare item performance for female and male students as well as racial/ethnic groups. Table 2.3 indicates which group was favored by the DIF, which means that the group performed better than expected on the item when controlling for performance on the test overall. Note that although DIF is statistical evidence that an item may be biased, approximately 5% of items are expected to be flagged even when there is truly no DIF. In general, DIF flagging rates are near or below the expected 5% when there is no DIF.

Table 2.3. ACT Test Items Exhibiting DIF based on 2025 Wisconsin Student Data

Subject	Reference Group	Focal Group	Number of Items	No DIF	DIF
English	Male	Female	75	75	0
	Non- EL	English learner (EL)	75	75	0
	White	African-American	75	75	0
	White	Asian	75	75	0
	White	Hispanic	75	75	0
	White	Two or more races	75	75	0
Mathematics	Male	Female	60	60	0
	Non-EL	English learner	60	60	0
	White	African-American	60	60	0
	White	Asian	60	60	0
	White	Hispanic	60	60	0
	White	Two or more races	60	60	0
Reading	Male	Female	40	40	0
	Non-EL	English learner	40	40	0
	White	African-American	40	40	0
	White	Asian	40	40	0
	White	Hispanic	40	40	0
	White	Two or more races	40	40	0
Science	Male	Female	40	40	0
	Non-EL	English Learner	40	40	0
	White	African-American	40	40	0
	White	Asian	40	40	0
	White	Hispanic	40	40	0
	White	Two or more races	40	40	0

Note: Information in this table can also be found in Appendix Table 12.

2.5 Test Development Procedures for the Writing Test

This section describes the procedures for developing essay prompts for the ACT writing test. These include many of the same steps used to develop the multiple-choice tests.

2.5.1 Prompt Writers

ACT writing prompts are produced by internal content specialists. ACT writing specialists have broad professional experience in secondary and postsecondary classrooms and in the field of writing assessment.

2.5.2 Prompt Construction

Prompts developed for the writing test provide topics with enough complexity and depth that examinees can write thoughtful and engaging essays. Topics are carefully chosen so that they are neither too vast nor too simplistic and do not require specialized prior knowledge. In constructing prompts, ACT writing specialists take into account that a student must be able to respond within the 40-minute time constraint of the test.

2.5.3 Content and Fairness Review of Prompts

After writing test prompts are developed and refined by ACT writing specialists, the prompts go through a rigorous review process with external experts. These fairness and bias experts carefully review each prompt to ensure that neither the language nor the content of a prompt will be offensive to a test taker and that no prompt will disadvantage any student from any geographic, socioeconomic, or cultural background. Reviewers also help ensure that prompts are accessible and engaging to students by evaluating prompt content in relation to student knowledge, experience, and interests.

2.5.4 Field Testing of Prompts

ACT conducts a special field test study periodically to evaluate new ACT writing prompts and to select those suitable for operational use. Students from across the United States—from rural and urban settings, small and large schools, and public and private schools—write responses to the new prompts, which are then read and scored by ACT-trained readers.

Prompts are evaluated from both content and statistical perspectives to ensure that scores (reported on a scale of 2 to 12) are comparable across different test forms and different administrations. In each field test study, anchor prompts and new prompts are administered to randomly equivalent groups of approximately 1,000 students per prompt.

Each student takes two prompts, and the order in which the prompts are taken is counterbalanced. Prompts are spiraled within classrooms so that, across all participating students, randomly equivalent groups of students take each prompt, with about half of the students taking a prompt first and the rest taking it second.

2.5.5 Review of Field Tests and Operational Administration

Once scoring of the new writing test prompts has been completed, the prompts are statistically analyzed to judge their acceptability. ACT applies the acceptability criteria after examining the relationships among scores on newly field-tested prompts and older (anchor) prompts. Specifically, the 2-to-12 score distributions should align, and there should be students scoring at the top of the score scale. Also, equating results should show that equating errors are within expected ranges at all score points, and the raw-to-scale score conversion tables, which are used to generate scores (from 1 to 36) that contribute to the ACT ELA score, exhibit desirable properties (see *Chapter 6.2* for more information about writing equating).

2.6 ACT Scores

This section briefly introduces the scores generated from student responses to the ACT test. *Chapter 5* provides additional information about these scores and ACT score reports. This section concludes with a summary of ACT policies concerning scoring appeals and inquiries.

2.6.1 ACT Scale Scores

For each test section on the ACT (English, mathematics, reading, and science), the raw scores (number of correct multiple-choice responses) are converted to scale scores ranging from 1 to 36. The Composite score is the average of the four content test scale scores rounded to the nearest whole number (fractions of 0.5 or greater round up). The minimum Composite score is 1; the maximum is 36. See *Chapter 6* for more details about the creation and maintenance of the 1-to-36 ACT scales.

If the student took the writing test, the student's essay is read and scored independently by two trained raters, one of which may be CRASE+®, ACT's automated essay scoring engine. Essays are scored analytically—that is, on the basis of traits in the essay that correspond to four domains of writing identified in the scoring rubric: Ideas and Analysis, Development and Support, Organization, and Language Use and Conventions. Each reader rates an essay on a scale ranging from 1 to 6 for each of the four domains. The sum of the readers' ratings for each domain is the domain score, reported on a scale ranging from 2 to 12. The subject-level writing test score, also 2 to 12, is the rounded average of the four domain scores. Writing scores are converted to a 1-to-36 scale only for the purpose of calculating the ELA score; the 1-to-36 writing scores are not reported.

2.6.2 STEM and ELA Scores

Since fall 2015, ACT has reported a Science, Technology, Engineering, and Math (STEM) score, which is calculated as the average of the 1-to-36 mathematics and science scale scores rounded to the nearest integer (fractions of 0.5 or greater round up). Only students who receive scores on the mathematics and science tests receive an ACT STEM score.

In fall 2015, ACT also began reporting a combined ELA score. The ACT ELA score is the rounded average of the English score, the reading score, and the 1-to-36 writing scale score. Only students who take all three of these tests can receive an ELA score. For the calculation of

ELA scores, the sum of the writing domain scores is converted to a scale of 1 to 36. However, this 1-to-36 writing scale score is not reported independently. Procedures for obtaining the 1-to-36 writing scale scores are described in *Chapter 6*.

2.6.3 Reporting Category Scores and Readiness Ranges

English, mathematics, reading, and science items align with reporting categories linked to the ACT College and Career Readiness Standards and other standards that target college and career readiness. There are three reporting categories each for English, reading, and science and eight for mathematics. Students receive a score in each reporting category, and score reports show corresponding Readiness Ranges, which indicate the range of scores expected of students who met or exceeded the ACT College Readiness Benchmark in that content area. The ACT Readiness Ranges appear on the Student Score Report and the High School Score Report. The combination of reporting category scores and the ACT Readiness Ranges provides educators and students with information that more clearly shows where students require the most assistance. Descriptions of the reporting categories are provided in *Chapter 3*. An explanation of how the readiness ranges are developed is provided in *section 6.1.4*.

2.6.4 Understanding Complex Texts Indicator

ACT test score reports include an Understanding Complex Texts indicator to show whether students understand the central meaning of complex texts at a level that is needed to succeed in college courses with higher reading demands. This indicator is based on scores from a subset of items on the reading test. These items measure a more global comprehension of the passages instead of sentence- or word-level understanding. Student performance on these items is divided into three performance levels: Below Proficient, Proficient, and Above Proficient.

2.6.5 Scoring Appeals and Inquiries

Electronic scanning devices are used to score the four multiple-choice tests of the ACT, thus minimizing the potential for scoring errors. If a student believes that a scoring error has been made, ACT hand-scores the answer document (for a fee) upon receipt of a written request from the student. Strict confidentiality of each student's record is maintained. In the spring 2025 Wisconsin state-sponsored administration of the ACT, 108 student scores were cancelled due to misadministration (N=52) or aberrant student behaviors (N=56).

If a student believes that a writing test essay has been incorrectly scored, that score may be appealed. ACT will verify (for a fee) that the essay was scored by at least two independent, qualified readers—one of which may have been CRASE+-- and by a third reader in the event that the two scores differed by more than one point in any domain. ACT will also verify that the essay was properly captured and displayed to readers. If errors are discovered during score verification, ACT will rescore the essay and refund the score verification fee.

For certain test dates (found online at *www.act.org*), examinees may obtain (for a fee) a copy of the test items used in determining their scores, a list of the correct answers, a list of their answers, and a table to convert raw scores to the reported scale scores. (For an additional fee, a student may also obtain a copy of his or her answer document.) These materials are available

only to students who test during regular administrations of the ACT on specified national test dates. If for any reason ACT must replace the test form scheduled for use at a test center, this offer is withdrawn and the student's fee for this optional service is refunded.

ACT reserves the right to cancel test scores when there is reason to believe the scores are invalid. Cases of irregularities in the test administration process—falsifying one's identity, impersonating another examinee, unusual similarities in answers of examinees at the same test center, examinee misconduct, or other indicators that the test scores may not accurately reflect the examinee's level of educational achievement—may result in ACT's canceling the test scores. For a detailed description of how ACT handles score cancelations, refer to ACT's Terms and Conditions of Registration (*www.act.org/the-act/terms*).

Chapter 3
Content Specifications

3.1 Overview

The ACT® test is constructed to meet specifications for content balance within the assessment domains. The content specifications define ranges for the number of items in each content category and at each level of cognitive complexity. The content specifications may also set test-specific requirements for the number of passages, distribution of passage genres, passage and item word counts, and diverse representation in passages in terms of gender, ethnicity, region, and community type (urban or rural). These content blueprints ensure that the knowledge and skills in the content domains are sampled consistently across test forms. The following chapter describes the assessment domain and content blueprint for each of the four multiple-choice ACT tests and the optional writing test.

3.2 English Test

3.2.1 Description of the English Test

The ACT English test is a 75-item, 45-minute test that puts the student in the position of a writer who is revising and editing a text. The test measures a student's understanding of the conventions of standard written English (grammar, usage, and mechanics), production of writing (topic development, organization, unity, and cohesion), and knowledge of language (word choice, style, and tone). The test consists of five passages, each accompanied by a sequence of multiple-choice test items. Different passage types are employed to provide a variety of rhetorical situations. Students must use the rich context of the passages to make editorial choices, demonstrating their understanding of writing strategies and conventions. Passages are chosen not only for their appropriateness in assessing writing and language skills but also to reflect students' interests and experiences. Spelling and the rote recall of grammar rules are not tested.

Some items refer to underlined or highlighted portions of the passage and offer several alternatives to the designated portion. These items often include making no change to the designated portion of the passage as one of the possible responses. Some items are identified by a number in a box or by a highlighted asterisk. These items ask about a section of the passage or about the passage as a whole. Some items appear at the end of the item set and are accompanied by instructions noting that the questions are about the passage as a whole. The student must decide which choice best answers each question.

Cognitive Complexity and Depth of Knowledge (DOK)

DOK (Webb, 2002) is a rough-grained, judgment-based measure of a test item's cognitive complexity that is used in many educational contexts. The ACT English test assesses skills that vary in cognitive complexity using items at DOK Levels 1, 2, and 3. All English items are classified by ACT content experts according to the level descriptions in Table 3.1.

Table 3.1. DOK Level Descriptions for English

Depth of Knowledge Level	Description
DOK1	Requires the recall of information, such as a fact, term, definition, or simple procedure.
DOK2	Requires mental processing that goes beyond recalling or reproducing an answer. Students must make some decisions about how to approach a problem.
DOK3	Requires planning, thinking, explaining, justifying, using evidence, conjecturing, and postulating. The cognitive demands are complex and abstract.

3.2.2 English Scores and Reporting Categories

Four scores are reported for the ACT English test: a total test score based on all 75 items and three reporting category scores. The total test score is reported on the ACT English scale, which ranges from 1 to 36. That score is averaged with the reading and writing test scores to determine the ELA score (see *Chapter 5* for more information about the derivation of the ELA score). The three reporting categories associated with the English test are Production of Writing, Knowledge of Language, and Conventions of Standard English. These reporting categories are subdivided into six elements, each of which targets an aspect of effective writing. A brief description of the reporting categories is given below. ACT score reports provide the percentage of correctly answered items in each reporting category and a Readiness Range indicating the range of scores expected of students who meet the ACT College Readiness Benchmark for English (18).

Production of Writing

Students apply their understanding of the rhetorical purpose and focus of a piece of writing to develop a topic effectively. They use various strategies to achieve logical organization, topical unity, and cohesion.

Topic Development

Students demonstrate understanding and control of rhetorical aspects of texts by identifying the functions of parts of texts, determining whether a text or part of a text has accomplished a purpose, and evaluating the relevance of material in terms of a text's focus.

Organization, Unity, and Cohesion

Students use various strategies to ensure that a text is logically organized, flows smoothly, and has an effective introduction and conclusion.

Knowledge of Language

Students demonstrate effective language use by ensuring precision and concision in word choice and maintaining consistency in style and tone.

Conventions of Standard English

Students apply their understanding of the conventions of Standard English grammar, usage, and mechanics to revise and edit text.

Sentence Structure and Formation

Students apply an understanding of sentence structure and formation, including understanding the placement of modifiers and relationships between and among clauses.

Usage

Students edit text to conform to Standard English usage.

Punctuation

Students edit text to conform to Standard English punctuation.

3.2.3 English Test Blueprints

Table 3.2 shows the current target distribution of test items across reporting categories on each ACT English test form.

Table 3.2. Specification Ranges by Reporting Category for English

Reporting Category	Number of Items	Percentage of Test
Production of Writing	22–24	29–32%
Knowledge of Language	11–13	15–17%
Conventions of Standard English	39–41	52–55%
Total number of items	75	100%

3.3 Mathematics Test

3.3.1 Description of the Mathematics Test

The ACT mathematics test is a 60-item, 60-minute test that measures the whole of a student's mathematical development up through topics typically taught at the beginning of Grade 12 in U.S. schools, focusing on prerequisite knowledge and skills important for success in college mathematics courses and career training programs. The domain is divided into Preparing for Higher Mathematics (PHM) and Integrating Essential Skills (IES).

The mathematics construct requires making sense of problems and context; representing relationships mathematically; accessing appropriate mathematical knowledge from memory; incorporating given information; modeling; doing mathematical computations and manipulations; interpreting; applying reasoning skills; justifying; making decisions based on the mathematics; and appropriately managing the solution process. The test emphasizes quantitative reasoning and application over extensive computation or memorization of complex formulas. Items focus

on what students can do with the mathematics they have learned, which encompasses not only mathematical content but also mathematical practices.

Some degree of computational fluency is required. A calculator is encouraged but not required. Items are designed so that a sophisticated calculator does not provide a significant advantage over a four-function calculator. Items are also designed so that all problems can be done without a calculator in a reasonable amount of time.

Each item has five response options. The test contains problems ranging from easy to challenging in order to reliably report on readiness levels for students with different preparation.

The mathematics test may include up to two item sets. An item set first presents information, including text, graphs, or other stimulus material, and then follows that information with a set of two to five items that each draw upon the given information. Items in the set, and across the form in general, are chosen to be logically independent, meaning that getting the correct answer to one item does not depend upon getting the correct answer to another item.

Cognitive Complexity and Depth of Knowledge (DOK)

The ACT mathematics test assesses skills that vary in cognitive complexity using items at DOK Levels 1, 2, and 3. All mathematics items are classified by ACT content experts according to the level descriptions in Table 3.3.

Table 3.3. DOK Level Descriptions for Mathematics

Depth of Knowledge Level	Description
DOK1	Requires the recall of information, such as a fact, term, definition, or simple procedure. Requires students to demonstrate a rote response or perform a simple procedure.
DOK2	Requires mental processing that goes beyond recalling or reproducing an answer. Students must make some decisions about how to approach a problem.
DOK3	Requires planning, thinking, explaining, justifying, using evidence, conjecturing, and postulating. The cognitive demands are complex and abstract.

3.3.2 Mathematics Scores and Reporting Categories

Nine scores are reported for the ACT mathematics test: a total test score based on all 60 items and eight reporting category scores. The total test score is reported on the ACT mathematics scale, which ranges from 1 to 36. That score is averaged with the science score to determine the STEM score, which is related to success in postsecondary science, technology, engineering, and mathematics courses (see *Chapter 5* for more information about the derivation of the STEM score).

There are eight mathematics reporting categories designed to give more detail about a student's mathematical achievement. The additional reporting category scores show a pattern of strengths and weaknesses that can differ among students with the same mathematics test score. The test is first divided into Preparing for Higher Mathematics (PHM) and Integrating Essential Skills (IES) reporting categories. The PHM score is then divided into separate scores for Number & Quantity, Algebra, Functions, Geometry, and Statistics & Probability. A crosscutting reporting category, Modeling, draws upon items from all the other categories to give a measure of producing, interpreting, understanding, evaluating, and improving models. Table 3.4 shows the number of items that contribute to each reporting category score. Descriptions of each reporting category follow. ACT score reports provide the percentage of items in each reporting category answered correctly and a Readiness Range indicating the range of scores expected of students who meet the ACT College Readiness Benchmark for mathematics (22).

Preparing for Higher Mathematics

This reporting category captures the more recent mathematics that students are learning. This category is divided into the following five subcategories.

Number & Quantity

Students demonstrate an understanding of and fluency with rational numbers and the four basic operations, and they work with irrational numbers by manipulating rational numbers that are close. Students use properties of the real number system. Students show their knowledge of complex numbers, compute in this system, and work with the properties of complex numbers. Students use vectors and matrices and view them as number systems with properties, operations, and applications.

Algebra

Students use their understanding of linear equations to make sense of other kinds of equations and inequalities: what their graphs look like, how to solve them, and what kinds of applications they have for modeling. Students use expressions to solve problems, and they show an understanding of solving equations. Students demonstrate extended proficiency with equations by using quadratic, polynomial, rational, and radical equations as well as systems of equations. Students create expressions, equations, and inequalities to represent problems and constraints. Students see rational expressions as systems analogous to rational numbers, apply the binomial theorem, and solve simple matrix equations that represent systems of linear equations.

Functions

Understanding the general properties of functions equips students for problem-solving with new functions they create. Functions provide a framework for modeling real-world phenomena, and students interpret the characteristics of functions in the context of a problem. Students work with functions that have no equation and functions that follow the pattern of an equation. Students reason with particular families of functions—like linear, quadratic, and exponential—by looking at rates of change, algebraic properties, and connections to graphs and tables, and by applying

these functions in modeling situations. Students also work with a range of functions, like those defined in terms of square roots, cube roots, polynomials, exponentials, logarithms, and trigonometric relationships, as well as piecewise-defined functions.

Students have seen shifts in graphs due to parameter changes, but now they demonstrate a unified understanding of translations and scaling through forms such as $f(x - c)$, $f(x) + c$, $af(x)$, and $f(-ax)$. Students connect the trigonometry of right triangles to the unit circle to make trigonometric functions. They use these functions to model periodic behavior.

Students graph rational functions and demonstrate knowledge of asymptotes. They compose functions and use inverse functions to solve equations with more than one solution, in particular for trigonometric functions. They apply the algebraic properties of trigonometric functions, such as angle addition properties.

Geometry

Students show understanding of congruence and rigid motions, dilations, and similarity. They make geometric constructions, solve problems, and model with geometric objects. Students find values such as the area of a circle and the volume of cylinders, pyramids, and cones. Students demonstrate understanding of trigonometric ratios as functions of angles, and they solve right-triangle problems. In the coordinate plane, students derive conditions for parallel and perpendicular lines, split a line segment into pieces with a given ratio of lengths, find areas, and develop equations for circles and for parabolas.

Students use trigonometry to derive a formula for the area of a general triangle in terms of side lengths and the sine of an angle, and they apply the law of sines and law of cosines to answer questions about non-right triangles. They derive equations for ellipses and hyperbolas. Students show understanding of Cavalieri's principle when using formulas such as the formula for the volume of a sphere.

Statistics & Probability

Students demonstrate learning about the role of randomness in sample surveys, experiments, and observational studies. Students use data to estimate a population mean or proportion and make informal inferences based on their judgment of likelihood. They compare qualities of research reports based on data and use simulation data to make estimates and judgments.

Students demonstrate understanding of statistical independence. They relate the sample space to events defined in terms of "and," "or," and "not," and they calculate probabilities using empirical results, independence assumptions, and the ideas of conditional probability. Students understand the multiplicative rule for conditional probability and apply permutations and combinations as tools for counting. They model a sample space with a random variable by giving a numerical value to each event. Students apply expected value and probability to help inform their decisions.

Integrating Essential Skills

This reporting category focuses on whether students can put together knowledge and skills to solve problems of moderate to high complexity. Topics include rate and percentage; proportional reasoning; area, surface area, and volume; quantities and units; expressing numbers in different ways; using expressions to represent quantities and equations to capture relationships; rational exponents; the basics of functions; function notation; sequences as functions; transformations, congruence, symmetry, and rigid motions; data analysis and representation; measures of center and spread; normal distribution; associations between two variables; two-way tables; scatterplots; linear models; correlation; and model fit.

In addition to learning more content over time, students should grow in sophistication, accumulating and applying skills in higher-order contexts. Therefore, the ACT mathematics test requires students to solve problems of increasing complexity, combine skills in longer chains of steps, apply skills in more varied contexts, understand more connections, and increase fluency. To assess whether students have acquired such skills, the items in this reporting category are at least at DOK Level 2, with a significant portion at DOK Level 3. DOK is judged relative to well-prepared students in Grades 11–12.

Modeling

Modeling uses mathematics to represent, through a model, an analysis of an empirical situation. Models often help us predict or understand the actual. However, sometimes knowledge of the actual helps us understand the model, such as when addition is introduced to students as a model of combining two groups. The Modeling reporting category represents all items that involve producing, interpreting, understanding, evaluating, and improving models. Each modeling item is also counted in the other appropriate reporting categories above. Thus, the Modeling reporting category is an overall measure of how well a student uses modeling skills across mathematical topics.

3.3.3 Calculator Policy

Students are encouraged to bring a calculator they are familiar with and can use fluently. Most four-function, scientific, or graphing calculators are permitted. Built-in computer algebra systems are not allowed because they could interfere with the construct, specifically understanding and implementing solutions to various types of equations and inequalities. Students must remove certain kinds of programs from their calculators. Some calculator features are not allowed or must be turned off for security reasons or to avoid disruptions during testing. Current details are available on the *ACT website*. A graphing calculator is available in the online testing application in the mathematics subject-level test.

3.3.4 Mathematics Test Blueprints

Table 3.4 shows the current target distribution of test items across reporting categories on each ACT mathematics test form. Test construction also takes into account coverage and variety within each of the categories. As explained above, PHM represents newer topics, and the assessment includes items representing DOK Levels 1, 2, and 3. IES represents topics that

should be very familiar, and what is important for college readiness is putting these familiar skills to work in higher-complexity tasks (DOK2 and DOK3).

Table 3.4. Specification Ranges by Reporting Category for Mathematics

Reporting Category	Number of Items	Percentage of Test
Preparing for Higher Mathematics	34–36	57–60%
Number & Quantity	5–7	8–12%
Algebra	7–9	12–15%
Functions	7–9	12–15%
Geometry	7–9	12–15%
Statistics & Probability	5–7	8–12%
Integrating Essential Skills	24–26	40–43%
Modeling	≥12	≥20%
Total number of items	60	100%

***Notes*:** Each item reported in Modeling is also reported in either Preparing for Higher Mathematics (and the appropriate subcategory) or in Integrating Essential Skills.

3.4 Reading Test

3.4.1 Description of the Reading Test

The ACT reading test is a 40-item, 35-minute test that measures a student's ability to read closely, reason about texts using evidence, and integrate information from multiple sources. The test comprises four passage units, three of which contain one long prose passage and one of which contains two shorter prose passages. Passages in the reading test include both literary narratives and informational texts from the humanities, natural sciences, and social sciences. Informational passages may include mixed-information formats—that is, visual and quantitative elements that accompany the text and contain additional information related to the passage topic. Passages are representative of the kinds of texts commonly encountered in high school and first-year college courses. Each passage is preceded by a heading that identifies the passage type (Literary Narrative or Informational), names the author, and may include a brief note that helps in understanding the passage by providing important background information.

Each passage unit includes a set of 9–11 multiple-choice test items. The items focus on the mutually supportive skills that readers apply when studying written materials across a range of subject areas. Specifically, items ask students to determine main ideas; locate and interpret significant details; understand sequences of events; make comparisons; comprehend cause-effect relationships; determine the meaning of context-dependent words, phrases, and statements; draw generalizations; analyze the author's or narrator's voice or method; analyze claims and evidence in arguments; and integrate information from multiple related texts and from different formats (e.g., graphs, diagrams, tables). Items do not test the rote recall of facts from outside the passage or rules of formal logic, nor do they contain questions about vocabulary that can be answered without referring to the passage context.

Cognitive Complexity and Depth of Knowledge (DOK)

The ACT reading test assesses skills that vary in cognitive complexity using items at DOK Levels 1, 2, and 3. All reading items are classified by ACT content experts according to the level descriptions in Table 3.5.

Table 3.5. DOK Level Descriptions for Reading

Depth of Knowledge Level	Description
DOK1	Requires the recall of information, such as a fact, term, definition, or simple procedure. Requires students to demonstrate a rote response or perform a simple procedure.
DOK2	Requires mental processing that goes beyond recalling or reproducing an answer. Students must make some decisions about how to approach a problem.
DOK3	Requires planning, thinking, explaining, justifying, using evidence, conjecturing, and postulating. The cognitive demands are complex and abstract.

3.4.2 Reading Scores and Reporting Categories

Four scores are reported for the ACT reading test: a total test score based on all 40 items, and three reporting category scores based on specific knowledge and skills. Score reports also include an Understanding Complex Texts indicator, which indicates proficiency (below, proficient, or above) in understanding the central meaning of complex texts at a level that is needed to succeed in college courses with high reading demand. The total test score is reported on the ACT reading scale, which ranges from 1 to 36. That score is averaged with the English and writing test scores to determine the ELA score ((see Chapter 5 for more information about the derivation of the ELA score). The three reporting categories addressed in the reading test are Key Ideas & Details, Craft & Structure, and Integration of Knowledge & Ideas. ACT score reports provide the percentage of items in each reporting category answered correctly and a Readiness Range indicating the range of scores expected of students who meet the ACT College Readiness Benchmark for reading (22).

Key Ideas & Details

Students read texts closely to determine central ideas and themes, summarize information and ideas accurately, understand relationships (including sequential, comparative, and cause-effect), and draw logical inferences and conclusions.

Craft & Structure

Students determine word and phrase meanings, analyze how an author uses word choice to achieve a rhetorical effect, analyze text structure, understand authorial purpose and perspective, and analyze points of view. They interpret the rhetorical effects of authorial decisions and differentiate between various perspectives and sources of information.

Integration of Knowledge & Ideas

Students understand authors' claims, differentiate between facts and opinions, and use evidence to make connections between different texts. Some items will require students to analyze how authors construct arguments, evaluating reasoning and evidence from various sources. Items in this category may ask students to interpret information presented in visual and quantitative formats (e.g., graphs, diagrams, or tables) and integrate this information with that in the passage text (see *Section 3.4.4* for more information).

3.4.3 Reading Test Blueprints

Table 3.6 shows the current target distribution of test items across reporting categories on each ACT reading test form.

Table 3.6. Specification Ranges by Reporting Category for Reading

Reporting Category	Number of Items	Percentage of Test
Key Ideas & Details	21–24	53–60%
Craft & Structure	10–12	25–30%
Integration of Knowledge & Ideas	6–9	15–23%
Total number of items	40	100%

3.4.4 Visual and Quantitative Information

To improve alignment between the ACT reading test and state English language arts content standards, ACT began developing reading passages and items that require students to interpret visual and quantitative information (VQI). ACT's plan is for one of the four reading passages on each test form to include VQI and for two associated items to measure students' skills related to interpreting and solving problems with VQI. This new type of content is also referred to as a mixed information format. Although the skills for comprehending this type of reading content are included in states' English language arts reading standards and belong to the content domain of the assessment, the skills measured by such items are different in nature from those measured by other ACT reading items. Thus, it was important to evaluate whether the addition of VQI passages and items had any notable impacts on the psychometric properties of the ACT reading test. To date, ACT has conducted two sets of analyses on data from VQI units, and these are summarized below. Both analyses indicated that VQI content was statistically indistinguishable from non-VQI reading content. That is, VQI items were not unusual in terms of difficulty, discrimination, differential item functioning (DIF), or their contribution to measurement precision (reliability). As more data become available, ACT can conduct analyses to determine the extent to which VQI items measure a slightly different construct than non-VQI reading items.

In 2019, ACT reworked five preexisting reading units. This involved adding VQI content to the passages, shortening other parts of the passages, and developing three VQI items for each revised passage. The VQI units were spiraled into the February 2020 field test booklets, which were appended to ACT test booklets like other newly developed content (as the "fifth test"). Following the February 2020 administration, ACT conducted psychometric analyses to examine

whether the VQI units functioned like non-VQI reading units. The VQI items had a range of difficulties (proportions correct) between 0.40 and 0.75, which was well within the typical and acceptable range for reading items. With point-biserial correlations ranging from approximately 0.38 to 0.52, the VQI items were also found to be acceptably discriminating between examinees of lower and higher ability. Internal-consistency reliability (coefficient alpha) was calculated for the sets of 14–15 items associated with each VQI passage. Those reliability coefficients ranged from 0.59 to 0.82, which was similar to the range of 0.64 to 0.84 for non-VQI units. Item response theory (IRT) was employed to evaluate model-data fit for VQI units (i.e., the degree to which the observed data for an item correspond to expectations), and results indicated total scores on the VQI units were well aligned with expectations based on the measurement model. Finally, the VQI items were examined for evidence of possible gender bias. A DIF analysis revealed that male and female examinees were equally likely to respond correctly to VQI items when controlling for overall achievement.

Of the VQI units that were field tested in February 2020, three units were included in new reading forms that were equated in February 2021 (note that the reading test blueprint did not change—the VQI unit took the place of a non-VQI informational passage and its items). That is, three forms with VQI units were spiraled with other new forms (and an anchor form) in the February 2021 ACT administration to determine the relationship between number correct (raw) scores and 1–36 scale scores. This was the first time VQI units were administered operationally. Following that administration, the six VQI items (three passages with two items each) were again examined. Again, the VQI items did not stand out among the reading items on those forms. The VQI items had proportions correct of 0.66, 0.57, 0.87, 0.47, 0.57, and 0.53, and they had point-biserial correlations of 0.47, 0.30, 0.29, 0.21, 0.33, and 0.36. As is typical, the operational proportions correct were slightly higher than the field test values reported above. As for potential item bias, none of the VQI items were flagged for DIF when comparing genders or racial/ethnic groups. Considering that the six VQI items were statistically indistinguishable from the non-VQI reading items, it was not surprising that the 40-item reading forms in which they were embedded had properties similar to those of the forms without VQI units. For example, the average proportions correct for the three VQI forms were 0.55, 0.56, and 0.58, and the range for the other forms was 0.55 to 0.59. The coefficient alphas for the VQI forms were 0.87, 0.87, and 0.88, and the range for the other forms was 0.86 to 0.89.

Based on analyses to date, ACT is confident that VQI units will continue contributing to the reliable measurement of reading skills. For monitoring, ACT will periodically analyze data from VQI units. When operational data from more VQI items become available, future analyses will include correlations between VQI items and other reading items to gauge the extent to which VQI items measure a slightly different construct.

3.5 Science Test

3.5.1 Description of the Science Test

The ACT science test is a 40-item, 35-minute test that measures the interpretation, analysis, evaluation, reasoning, and problem-solving skills required in the natural sciences. The content

of the science test is drawn from the following content areas, which are all represented on the test: biology, chemistry, physics, and Earth and space science.

Students are assumed to have a minimum of two years of high school introductory science, which ACT's National Curriculum Survey has identified as typically one year of biology and one year of physical science or Earth science. Thus, it is expected that students have acquired the introductory content of biology, physical science, and Earth science, are familiar with the nature of scientific inquiry, and have been exposed to laboratory investigation.

The test presents several sets of scientific information, each followed by a number of multiple-choice test items. The scientific information is conveyed in one of three formats: data representation (scientific graphs, tables, and diagrams), research summaries (descriptions of one or more related experiments), or conflicting viewpoints (two or more brief theoretical models that address the same scientific phenomenon but conflict with one another).

The test assesses and reports on science knowledge, skills, and practices across three domains: Interpretation of Data; Scientific Investigation; and Evaluation of Models, Inferences & Experimental Results. The knowledge and skills encompassed in each domain were derived from decades of ACT's empirical data and research on college and career readiness in science. The domains and their skills link with quantitatively determined score ranges for the ACT science test and the ACT College Readiness Benchmark in science, which is predictive of success in science courses at the postsecondary level.

In addition, some of the ACT science items require students to have discipline-specific content knowledge (e.g., knowledge specific to an introductory high school physical science or biology course), but all of the items focus on scientific processes and critical thinking skills.

Cognitive Complexity and Depth of Knowledge

The ACT science test assesses skills and practices that vary in cognitive complexity using items at DOK Levels 1, 2, and 3, with almost all the items being at DOK Levels 2 and 3. ACT science experts have worked with several Webb-based systems adapted for science, but none of those systems quite capture the different dimensions associated with items focused on science skills and practices. Even so, all science items are classified by ACT content experts according to the level descriptions in Table 3.7.

Table 3.7. DOK Level Descriptions for Science

Depth of Knowledge Level	Description
DOK1	Requires locating, recalling, and/or reproducing information.
DOK2	Requires processing presented information and applying skills and concepts. Students typically must process one or two cognitive steps.
DOK3	Requires use of higher-order thinking, such as analysis and evaluation, and often requires using evidence to justify reasoning. Students must typically process multiple cognitive steps, and the overall tasks tend to be complex and abstract.

3.5.2 Science Scores and Reporting Categories

Four scores are reported for the ACT science test: a total test score based on all 40 items and three reporting category scores based on different domains of scientific knowledge, skills, and practices. The total test score is reported on the ACT science scale, which ranges from 1 to 36. That score is averaged with the mathematics score to determine the STEM score, which is related to success in postsecondary science, technology, engineering, and mathematics courses (see Chapter 5 for more information about the derivation of the STEM score). The three reporting categories addressed in the science test are Interpretation of Data; Scientific Investigation; and Evaluation of Models, Inferences & Experimental Results. A description of each reporting category is provided below. ACT score reports provide the percentage of items in each reporting category answered correctly and a Readiness Range indicating the range of scores expected of students who meet the ACT College Readiness Benchmark for science (23).

Interpretation of Data

Students manipulate and analyze scientific data presented in tables, graphs, and diagrams (e.g., recognize trends in data, translate tabular data into graphs, interpolate and extrapolate, and reason mathematically).

Scientific Investigation

Students understand experimental tools, procedures, and design (e.g., identify variables and controls) and compare, extend, and modify experiments (e.g., predict the results of additional trials).

Evaluation of Models, Inferences & Experimental Results

Students judge the validity of scientific information and formulate conclusions and predictions based on that information (e.g., determine which explanation for a scientific phenomenon is supported by new findings).

3.5.3 Science Test Blueprints

Table 3.8 shows the current target distribution of test items across reporting categories on each ACT science test form. Table 3.9 shows the current target distribution of test items across science content areas on each ACT science test form.

Table 3.8. Specification Ranges by Reporting Category for Science

Reporting Category	Number of Items	Percentage of Test
Interpretation of Data	16–20	40–50%
Scientific Investigation	8–12	20–30%
Evaluation of Models, Inferences & Experimental Results	10–14	25–35%
Total number of items	40	100%

Table 3.9. Specification Ranges by Science Content Area

Science Content Area	Number of Passages	Number of Items	Percentage of Test
Biology	2	11–15	28–38%
Chemistry	1–2	5–15	13–38%
Physics	1–2	5–15	13–38%
Earth and Space Science	1–2	5–15	13–38%
Total	6	40	100%

3.6 Writing Test

3.6.1 Description of the Writing Test

The ACT writing test is an optional 40-minute essay test that measures students' writing skills—specifically those skills emphasized in high school English classes and entry-level college composition courses. Scores from the writing test indicate students' ability to think critically about an issue, consider different perspectives on it, and compose an effective argumentative essay.

The test consists of one writing prompt that describes a complex issue and provides three different perspectives on the issue. Students are asked to read the prompt and write an essay in which they develop their own perspective on the issue. The essay must analyze the relationship between their own perspective and one or more other perspectives. Students may adopt one of the perspectives given in the prompt as their own, or they may introduce one that is completely different from those given. Their score will not be affected by the point of view they take on the issue.

Cognitive Complexity and Depth of Knowledge (DOK)

The cognitive complexity of the writing test essay task is classified as DOK 3 (Table 3.10).

Table 3.10. DOK Level Description for Writing

Depth of Knowledge Level	Description
DOK3	Requires planning, thinking, explaining, justifying, using evidence, conjecturing, and postulating.

3.6.2 Writing Scores and Domains

Students who take the optional writing test receive five scores: a single subject-level writing score and four domain scores. The overall writing score is reported on the ACT writing scale,

which ranges from 2 to 12.[1] Taking the writing test does not affect the student's section test scores or Composite score. However, a writing test score, along with the overall English and reading test scores, is needed to produce the ELA score. The overall writing score (after it has been converted to a 1–36 scale) is averaged with the English and reading test scores to determine the ELA score (see Chapter 5or more information about the derivation of the ELA score).

The four writing domains are Ideas & Analysis, Development & Support, Organization, and Language Use & Conventions. A brief description of the writing domains is given below. Scores on the four domains are each reported on a 2–12 scale, and the overall writing score is the rounded average of the four domain scores. The domain scores are based on an analytic scoring rubric, and two trained raters score each essay on a scale of 1 to 6 in each of the four domains. If the ratings disagree by more than one point, a third rater evaluates the essay and resolves the discrepancy (see Chapter 5 for more information about writing performance scoring and the analytic scoring rubric).

Ideas & Analysis

Scores in this domain reflect the ability to generate productive ideas and engage critically with multiple perspectives on the given issue. Proficient writers understand the issue they are invited to address, the purpose for writing, and the audience. They generate ideas that respond to the situation.

Development & Support

Scores in this domain reflect the ability to discuss ideas, offer rationale, and strengthen an argument. Proficient writers explain and explore their ideas, discuss implications, and illustrate through examples. They help the rater understand their thinking about the issue.

Organization

Scores in this domain reflect the ability to organize ideas with clarity and purpose. Organizational choices are integral to effective writing. Proficient writers arrange their essay in a way that clearly shows the relationships among ideas, and they guide the reader through their discussion.

[1] Students who took the writing test between September 2015 and June 2016 received subject-level writing scores reported on a 1–36 scale rather than subject-level scores reported on the current 2–12 scale. It should also be noted that the current 2–12 subject-level writing scores are not comparable to the 2–12 scores from the former writing test (June 2015 and before). Although both tests measure a student's ability to write an effective argumentative essay, the current test has a new design. Moreover, the current test is scored with an analytic rubric, whereas the former writing test was scored with a holistic six-point rubric. The score on the former test was the sum of the two raters' 1–6 scores rather than the rounded average of four 2–12 domain scores.

Language Use & Conventions

Scores in this domain reflect the ability to use written language to clearly convey ideas. Proficient writers make use of the conventions of grammar, syntax, word usage, and mechanics. They are also aware of their audience and adjust the style and tone of their writing to communicate effectively.

Chapter 4
Test Administration, Test Security, and Accessibility and Accommodations

4.1 Test Administration Overview

The ACT® test must be administered in a standardized manner to ensure a fair and equitable testing environment for all examinees. Testing staff must strictly adhere to ACT policies and procedures during test administrations. This chapter provides a brief description of the processes used to administer the ACT in both paper and online formats.

4.1.1 Administration Windows

The state-sponsored ACT administrations are available to students on predetermined test dates. For ACT State and District testing, WDPI chooses from predetermined windows during the spring administration. WDPI selected the following dates for Spring 2025:

Testing Window	Online	Standard Paper	Accommodated Paper
1	March 11-14 & March 17-21	March 11	March 11-14 & March 17-21
2	March 25-28, March 31 & April 1-4	March 25	March 25-28, March 31 & April 1-4
3	April 8-11 & April 14-18	April 8	April 8-11 & April 14-18

4.1.2 Testing Modes

State and District testing sites have the option of administering the test on paper or online. The ACT administered online is the same test as the paper version but presented in an online delivery format. Online testing of the ACT is designed to provide test access over a short period of time and to accommodate makeup and emergency situations. Online administration of the ACT follows the administration guidelines established for paper testing, where appropriate. WDPI allows school districts to be able to choose whether to administer the ACT assessment on paper or online in order to best meet the needs of the school districts. Information about the comparability between these modes may be found in *Chapter 6*. In addition to standard formats, ACT offers accommodations and English learner (EL) supports for examinees approved for these accessibility supports.

4.1.3 Testing Locations

Wisconsin students participating in the state-sponsored administration were able to take the ACT assessment on school campuses during regularly scheduled school hours. State-sponsored school testing sites have the option of administering the test on paper or online.

4.1.4 Policies and Procedures

Administration Manuals

For both paper and online administrations, ACT provides Wisconsin schools and districts with a variety of documentation to support standardized administration of the test. The administration manuals provide detailed directions for selecting staff, maintaining test security, and administering tests in a standardized manner. The manuals cover topics such as:

- policies and procedures to follow before, during, and after testing;
- staffing levels, responsibilities of testing staff and staff training;
- prohibited behaviors;
- handling and documenting testing irregularities;
- documentation to be submitted to ACT after testing; and
- procedures for returning test materials to ACT.

All testing staff must read the documentation before test day and adhere to standardized procedures.

Staffing

Schools are responsible for providing both the facilities and testing staff (test coordinator, room supervisors and proctors).

All testing staff are required to administer and supervise the ACT in a nondiscriminatory manner and in accordance with all applicable laws, including the Americans with Disabilities Act.

Training Staff

For standardized testing to occur successfully, all staff must understand ACT policies and procedures and their own responsibilities for implementing them. It is critical that the same procedures are followed at every site. The test coordinator is responsible for providing testing staff with the proper manuals and training prior to test day.

All testing staff, both new and experienced, must attend a training session conducted by the test coordinator before test day to discuss policy, procedural, and logistical issues and ensure that everyone has a common understanding of what is to take place on test day.

A testing staff briefing session is required each test day morning, even with experienced staff. This is the time to ensure that all staff are present and make any necessary adjustments to staff assignments. The test coordinator should make sure that testing staff understand their responsibilities and should answer questions in a group setting so everyone has the same information at the same time.

4.2 Test Security

4.2.1 Prevention and Detection of Test Security Violations

To ensure the validity of ACT test score interpretations, the examinees, any individuals that have a role in administering the tests, and those who are otherwise involved in facilitating the testing process must strictly observe ACT's standardized testing policies and procedures. This includes the Test Security Principles set forth in ACT's administration manuals, which may be supplemented by ACT from time to time with additional communications to examinees and testing staff.

ACT's test security requirements are designed to ensure that examinees have equal opportunities to demonstrate their academic achievement and skills, that examinees who do their own work are not unfairly disadvantaged by examinees who do not, and that scores reported for each examinee have valid interpretations. Strict observation of the test security requirements is necessary to safeguard validity.

Testing staff must protect the confidentiality of the ACT test items and responses. Testing staff should be aware of their responsibilities and be competent to undertake their roles, including understanding ACT's test administration policies and procedures and acknowledging and avoiding conflicts of interest in their roles as test administrators for the ACT.

Testing staff must be alert to activities that can compromise the fairness of the test and the validity of score interpretations. Such activities include, but are not limited to, cheating and questionable test-taking behavior (such as copying answers or using prohibited electronic devices during testing), accessing questions prior to the test, taking photos or making copies of test questions or test materials, posting test questions on the Internet, and test proctor or test administrator misconduct (such as providing questions or answers to examinees or permitting them to engage in prohibited conduct during testing).

In addition to these security-related administration protocols, ACT engages in additional test security practices designed to protect ACT test content and the validity of score interpretations. These practices include (a) the use of a reporting hotline to ACT through which individuals can anonymously report information about misconduct on an ACT test, (b) data forensics to detect and respond to possible misconduct, and (c) web monitoring to detect testing misconduct, possible unauthorized disclosure of secure ACT test content, and any other activity that might compromise the security of the ACT test or the validity of score interpretations.

4.2.2 Information Security

ACT's Information Security framework is based on the widely recognized ISO/IEC 27000 standard (International Organization for Standardization, 2018). This framework was selected because it covers a range of information security categories that comprehensively matches the broad perspective that ACT takes in safeguarding information assets. These 13 categories covered by the framework are followed by brief statements of their importance to ACT:

1. Information Security Program Management: This is overseen by the information security officer at ACT. The information security officer is responsible for providing guidance and direction to the organization to ensure compliance with all relevant security-related regulations and requirements. The program itself is designed to cover all security domains identified in the ISO 27001 standards and provides comprehensive oversight for information security at ACT.

2. Information Security Risk Management: The cornerstone of the ACT Information Security program is a risk assessment that conforms to the ISO 27005 standard. The identification, management, and mitigation of information security risks are managed using the Information Security Management System (ISMS) guidelines defined in the 27005 standard. ACT also makes use of the SP NIST 800-37 Risk Assessment, which complies with Federal Information Security Management Act (FISMA) security requirements for risk management (National Institute of Standards and Technology, 2017).

3. Information Security Policies and Standards: ACT established an Information Security policy to set direction and emphasize the importance of safeguarding information and data assets. Additional supporting policies, standards, and procedures have been developed to communicate requirements.

    a. ACT's Information Security policy and the Assessment Data Sharing procedures govern the handling of student data that is classified as confidential restricted. The policy states that confidential restricted information must meet the following guidelines:

        - Electronic information assets must only be stored on ACT-approved systems or media with appropriate access controls.

        - Only limited authorized users may have access to this information.

        - Physical records must be locked in drawers or cabinets while not being used.

    b. As a comprehensive control system to protect student data, ACT also has Access Management, Business Continuity Standard, Clear Desk/Clear Screen, End User Storage, External Authentication, Information Security Incident Management, Malware Protection, Mobile Device, Network Security Management, Payment Card Security, Secure Application Development, Secure System Configuration, Security Event Logging and Monitoring Standard, System Vulnerability and Patch Management, and Web Content Standard.

4. Information and Technology Compliance: The systems that store, maintain, and process information are designed to protect data security through all life cycle stages. The security considerations surrounding ACT's systems include measures such as

encryption, system security requirements, and logging and monitoring to verify that systems are operating within expected parameters.

5. Business Continuity and Disaster Recovery: ACT maintains a Business Continuity program designed to provide assurance that critical business operations will be maintained in the event of a disruption. An essential part of the program includes a cycle of planning, testing, and updating. Disaster recovery activities are prioritized by the criticality of systems and recovery times established by the business owners.

6. Security Training and Awareness: At ACT, information security is everyone's responsibility. All employees take part in annual information security awareness training on topics covered in the Information Security policy. Additionally, ACT has individuals within the organization who are responsible for the management, coordination, and implementation of specific information security objectives and who receive additional information security training.

7. Identity and Access Management: ACT addresses data integrity and confidentiality by policies and procedures that (a) limit access to individuals who have a business need to know the information and (b) verify the individuals' identities. Access to ACT systems and data requires authorization from the appropriate system owner. Active directory, file permissions, and virtual private network (VPN) remote access are administered by an Identity and Access management team that is part of the information security organization.

8. Information Security Monitoring: The foundation of ACT's Information Security program is reflected in the Information Security policy, which is presented and reinforced with training to all ACT employees. ACT is held accountable to following the Information Security program through internal assessments of the security control environment. Additionally, ACT works with independent third parties to provide assessment feedback.

9. Vulnerability and Threat Management: ACT has several mechanisms in place to identify vulnerabilities on networks, servers, and desktops. Monthly vulnerability scanning is performed by a qualified approved scanning vendor (ASV). ACT has always maintained a "compliant" status in accordance with Payment Card Industry Data Security Standards (PCI DSS) requirements. In addition to the scans performed for PCI compliance, ACT has a suite of vulnerability scanning tools, which are coordinated with a log management and event-monitoring tool to provide reporting and alerting.

10. Boundary Defense: ACT utilizes multiple intrusion-protection and -detection strategies, tools, processes, and devices to look for unusual attack mechanisms and to detect compromise of these systems. Network-based intrusion detection system (IDS) sensors are deployed on Internet and extranet demilitarized zone (DMZ) systems and networks, which provide alerts and procedures for review and response. Procedures include security review and approval of changes to configurations, semiannual firewall rule

review, and restrictions to deny communications with or limit data flow to known malicious IP addresses.

11. Endpoint Defenses: A variety of tools are utilized to ensure that a secure environment is maintained at the end-user device level. This includes segmentation within ACT's network, antivirus programs, and data-loss prevention programs. VPN is required for all remote access to ACT's network. Wireless access on ACT's campus requires authentication credentials, and ACT continuously scans for rogue access points.

12. Physical Security: Maintaining security on the premises where information assets reside is often considered the first line of defense in information security. ACT has implemented several security measures to ensure that physical locations and equipment used to house data are protected, including card-key access to all facilities and camera monitoring at all entry points.

13. Security Incident Response and Forensics: Planning for how to handle information security incidents is a critical component of ACT's Information Security program. Formal policy guidance outlines the response procedures, notification protocols, and escalation procedures. Forensics are performed at the direction of the information security officer. In the event of a declared incident, ACT maintains a subscription service with a third party specializing in computer forensics.

ACT's Information Security Incident Response Plan (ISIRP) brings needed resources together in an organized manner to deal with an incident classified as an adverse event related to the safety and security of ACT networks, computer systems, and data resources.

The adverse event could come in a variety of forms: (a) technical attacks (e.g., denial of service attack, malicious code attack, exploitation of a vulnerability), (b) unauthorized behavior (e.g., unauthorized access to ACT systems, inappropriate usage of data, loss of physical assets containing confidential or confidential restricted data), or (c) a combination of activities. The purpose of the plan is to outline specific steps to take in the event of any information security incident.

The ISIRP charters an ACT Information Security Incident Response Team (ISIRT) with providing a coordinated security incident response throughout ACT around the clock (i.e., 24/7). Information security management has the responsibility and authority to manage the ISIRT and implement necessary ISIRP actions and decisions during an incident.

4.3 Test Administration and Accessibility Levels of Support

The accessibility supports permitted during testing are designed to remove barriers to examinee access to the test yet still honor the constructs the tests measure. It is important to abide by all outlined requirements for administering these supports. Types of accessibility supports for the ACT include:

- universal supports
- designated supports
- English learner (EL) supports
- accommodations

4.3.1 Universal Supports

Universal supports are available to all students and do not require ACT approval. These supports are embedded into testing practices. Examples of universal supports include, but are not limited to, the following:

- test booklet used as scratch paper (paper testing only)
- calculator for the mathematics section
- examinees allowed to ask for clarification of verbal instructions
- examinees allowed to ask for general administration directions to be repeated
- browser zoom/magnification (online testing only)
- “mark an item for review” function (online testing only)

4.3.2 Designated Supports

Designated supports may be available to any examinee for whom a need has been identified, but the underlying condition may not rise to the level of a disability. Most of these supports require advance planning to deliver. Examples of designated supports include, but are not limited to, the following:

- wheelchair accessibility (test at a table instead of a desk)
- permission for food, drink, or medication in the testing room
- permission to use a cushion
- permission to use a chair to prop up a leg
- seating in the front or back of the room

4.3.3 English Learner Supports

English learner (EL) supports are available only for examinees in U.S. schools who are not proficient in English. EL supports should be identified by the educators responsible for selecting supports needed to access curriculum, instruction, and assessments because of limited English proficiency. EL supports must be authorized by ACT prior to use.

An examinee's English proficiency changes over time, so EL supports expire and must be reauthorized after the expiration date noted on the decision notification. Current English proficiency is measured by an English Language Proficiency assessment in the four language domains of Reading, Writing, Speaking, and Listening taken within the previous 12 months.

EL supports are limited to the following:

- ACT-authorized word-to-word bilingual dictionary or glossary
- translated written test directions, provided by ACT
- one and one-half time
- small group testing

4.3.4 Accommodations

Allowed accommodations are available to users who have a documented disability. The ACT requires examinees who use accommodations to have a formally documented need for as well as relevant knowledge of and familiarity with these supports. Accommodations must be requested and authorized in advance according to the ACT testing procedures. Appropriate documentation of the accommodation need must be provided prior to testing by the examinee or by a local governing educational authority. Accommodations are available only for examinees with disabilities as documented in an IEP, 504 Plan, or another accommodations/supports plan.

Accommodations are intended to reduce or eliminate the effects of an examinee's disability; however, accommodations should never reduce learning expectations by reducing the scope, complexity, or rigor of an assessment. Accommodations provided on the ACT must be generally consistent with those provided for instruction and assessment in the educational environment. There are some accommodations that may be used in the educational environment that are not allowed for the ACT because they affect the validity of the assessment results (see *Section 4.3.5*, Modifications). There may be consequences for the use of unallowed or unauthorized accommodations during the ACT.

To the extent possible, accommodations should adhere to the following principles:

- Accommodations enable examinees to participate more fully and fairly in instruction and to demonstrate their knowledge and skills on the ACT.
- Accommodations are based on an examinee's need rather than on the category of an examinee's disability.

- Accommodations are based on a documented need in the instructional and assessment setting and should not be provided for the purpose of giving the examinee an enhancement that could be viewed as an unfair advantage or to obtain a desired score.
- Accommodations for an examinee with disabilities are described and documented in the examinee's appropriate educational plan.
- Accommodations become part of the examinee's program of daily instructions as soon as possible after completion and approval of the educational plan.
- Accommodations are not introduced for the first time during the ACT test.
- Accommodations are monitored for effectiveness during daily instruction.

Examples of accommodations include, but are not limited to, the following:

- timing or scheduling supports (e.g., extra testing time, breaks as needed)
- audio supports (e.g., human reading a Reader's Script aloud, text-to-speech, screen reader software)
- response supports (e.g., scribe to record responses, computer for constructed-response items, speech-to-text software for the writing test)
- sign language interpreter for verbal instructions
- alternate formats (e.g., braille, large print)

4.3.5 Modifications

Modifications are supports that are sometimes used during instruction to aid learning but, when used in a testing situation, may provide assistance in a manner that alters what the test measures. Thus, these modifications prevent the same type of access to performance related to the measured construct when compared to the performance of examinees taking unmodified assessments. Because modifications alter the construct being tested, scores from modified assessments cannot be compared to scores from unmodified assessments. Modifications are not available for the ACT test.

For additional information on accessibility supports for the ACT, please refer to these sources:

- *Accessibility Supports Guide for the ACT–State and District Testing*
- *The ACT Knowledge Hub: ACT Test Accessibility and Accommodations (TAA) System Supports*

Chapter 5
Scoring and Reporting

5.1 Overview

The ACT® test is composed of four multiple-choice test sections—English, mathematics, reading, and science—and an optional writing test. Score reports are provided to individual students, their high schools, and the colleges of each student's choice. The contents of the student, high school, and college score reports are slightly different because they serve different purposes. The reports all contain scores indicating students' performance on each test section and detailed information about students' performance on specific areas within each section. Additional information is provided on the score reports to make it easier to interpret scores and to help with college and career planning.

The ACT scores and indicators were introduced in *Chapter 2*. This chapter provides more detailed information about the scores and indicators as well as the scoring process for the writing test. Subsequent parts of this chapter describe the information provided on the score reports to facilitate college and career planning.

5.2 Test Section, Composite, STEM, and ELA Scores

The ACT student, high school, and college reports describe students' overall performance on the test sections. This includes 1–36 scale scores on each section as well as the Composite score and two combined scores. The combined scores are the science, technology, engineering, and mathematics (STEM) score, which is a combination of the student's mathematics and science scores, and the English language arts (ELA) score, which is a combination of the student's English, reading, and writing scores. Providing these scores constitutes a major section of score reports. For example, Figure 5.1 shows what students view online through MyACT, and Figure 5.2 shows a sample of the score report sent to high schools. Standard errors of measurement (SEMs), the ACT College and Career Readiness Benchmarks, and national (U.S.) and state ranks are also reported to make it easier to interpret these scores.

Figure 5.1. Overall Score and Percentile Rank on a Sample Interactive Score Report on MyACT

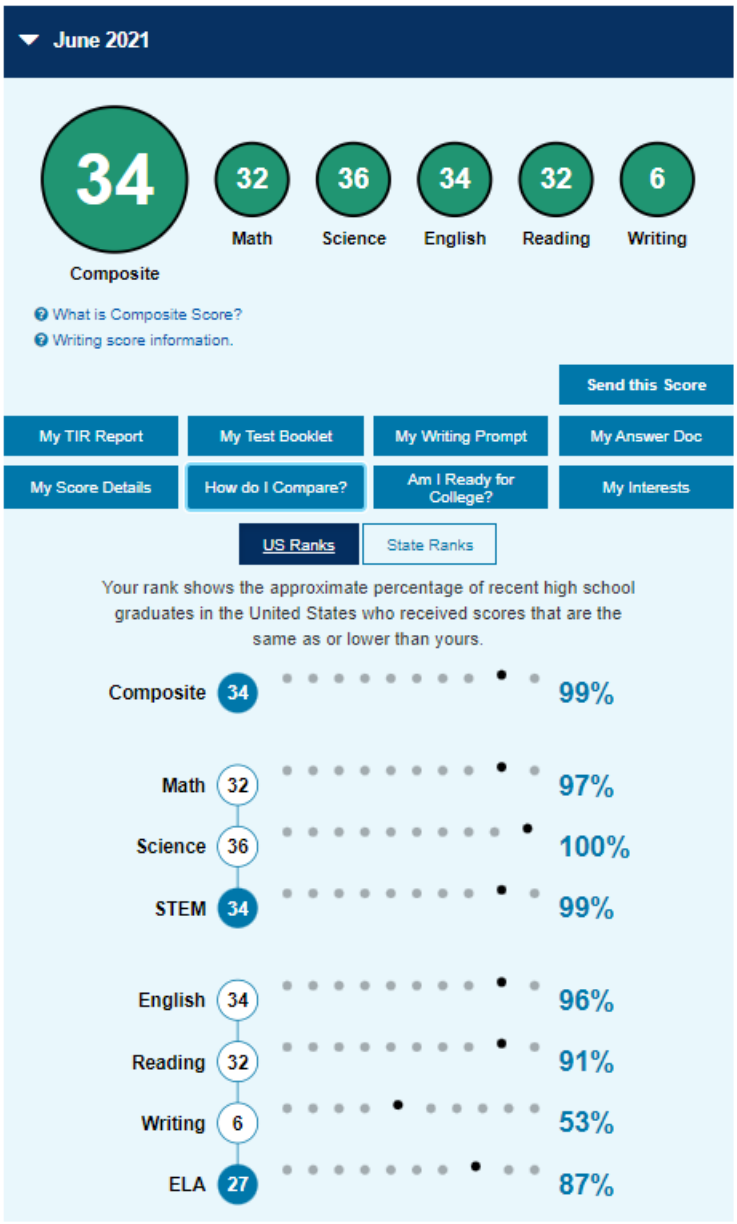
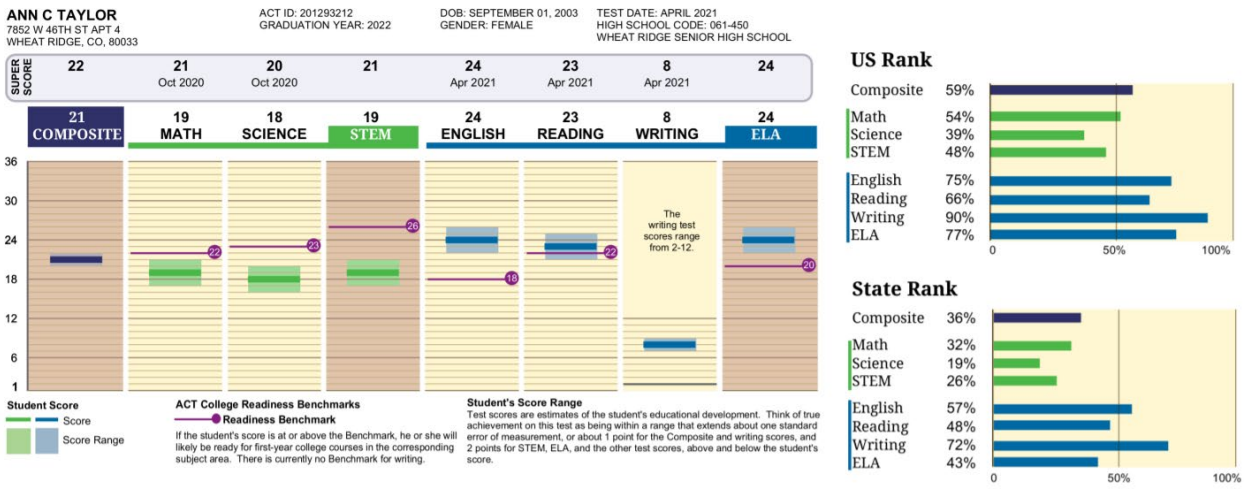


Figure 5.2. Overall Score and Percentile Rank on a Sample ACT High School Score Report



5.2.1 Test Section Scores

Multiple-Choice Tests

Test section scores are reported for the four multiple-choice tests. For each of the multiple-choice tests, the raw score is the number of test questions answered correctly. Raw scores are converted to scale scores through equating procedures to ensure that scores reported across test forms have consistent meaning. Scale scores range from 1 to 36 for each of the multiple-choice tests. Procedures for obtaining the 1–36 scale scores for the multiple-choice tests are described in *Chapter 6*.

Writing Test Scores

Student responses for the ACT writing test are scored by two trained raters (one of which may be CRASE+) on four writing domains: Ideas & Analysis, Development & Support, Organization, and Language Use & Conventions. Detailed descriptions of these domains are in *Chapter 3*. Using procedures described in *Section 5.2.2*, each rater assigns a score from 1 to 6 for each domain with an analytic rubric. Domain scores ranging from 2 to 12 are the sum of the two raters' scores. The writing test score is the average of the four domain scores rounded to the nearest whole number. The reported writing score ranges from 2 to 12.

5.2.2 Performance Scoring for the Writing Test

Various performance scoring processes and procedures are used for scoring the ACT writing test, such as range-finding, rater training and qualification, and rater monitoring. A scoring team composed of raters, scoring supervisors, scoring directors, and content specialists is responsible for these tasks. Team member roles and responsibilities are as follows:

- Raters complete a rigorous training course and must pass a qualifying test to participate in live scoring. All raters must have, at minimum, a 4-year degree from an accredited institution of higher education. Candidates with high school English teaching experience are preferred.

- Scoring supervisors are experienced expert raters. Each supervisor is responsible for a team of raters. Supervisors monitor the accuracy of raters, provide feedback to raters, and resolve discrepant scores.

- Scoring directors are performance scoring professionals. Directors are responsible for the overall management of scoring work, ensuring that scores are delivered on time and meet or exceed established quality parameters.

- Content specialists form a cross-functional team of assessment development, performance scoring, and education professionals with specific expertise and credentials in English language arts. Content specialists are responsible for range-finding, training development, and ongoing calibration.

Rater Training and Qualification

The range-finding process is the basis for developing scoring criteria validation and effective rater training materials. A panel of assessment and content experts meets to review a sample of student responses and ensure that content-specific criteria for each task accurately reflect and encompass the full range of student responses. Using consensus-scored responses, the panel builds exemplar “anchor” sets that will subsequently be used for rater training.

Developing these anchor sets of exemplar responses is the beginning of ACT’s rigorous training program. Anchor sets include multiple examples of responses at each score point and demonstrate a range of typical approaches to the assessment task. Each anchor response is fully annotated with scoring notes that link the student’s performance to the criteria described in the rubric (Table 5.1). In addition to anchor sets, ACT’s range-finding panels also develop practice and qualifying sets.

Rater candidates are introduced to the rubric and the writing prompt, and then they review these in concert with the prompt-specific anchor set. After becoming familiar with anchor responses, candidates are then given the opportunity to apply scores to multiple practice sets. Practice sets include a variety of responses, some of which are clearly aligned with particular score points and anchor responses, and others that require more detailed analysis to identify appropriate scores. Annotated feedback is provided at the conclusion of each practice set.

At the end of the training program, candidates are required to pass a qualifying test by perfectly matching a predetermined number of scores for at least two qualifying sets. Candidates who do not meet the qualifying standard are released from the scoring project.

A selected “baseline” prompt is used for rater training and qualification. All raters must participate in baseline training and pass the qualification test, which is administered at least twice annually. After qualifying, raters are introduced to additional writing prompts via prompt-specific anchor and practice sets, but raters do not need to re-qualify. The pool of raters is typically a diverse group in terms of age, ethnicity, and gender, although placement and retention of raters is based upon their qualifications and the quality and accuracy of their scoring.

Table 5.1. Writing Test Analytic Scoring Rubric

Score Point	Ideas & Analysis	Development & Support	Organization	Language Use & Conventions
Score 6: Responses at this score point demonstrate effective skill in writing an argumentative essay.	The writer generates an argument that critically engages with multiple perspectives on the given issue. The argument's thesis reflects nuance and precision in thought and purpose. The argument establishes and employs an insightful context for analysis of the issue and its perspectives. The analysis examines implications, complexities and tensions, and/or underlying values and assumptions.	Development of ideas and support for claims deepen insight and broaden context. An integrated line of skillful reasoning and illustration effectively conveys the significance of the argument. Qualifications and complications enrich and bolster ideas and analysis.	The response exhibits a skillful organizational strategy. The response is unified by a controlling idea or purpose, and a logical progression of ideas increases the effectiveness of the writer's argument. Transitions between and within paragraphs strengthen the relationships among ideas.	The use of language enhances the argument. Word choice is skillful and precise. Sentence structures are consistently varied and clear. Stylistic and register choices, including voice and tone, are strategic and effective. While a few minor errors in grammar, usage, and mechanics may be present, they do not impede understanding.
Score 5: Responses at this score point demonstrate well-developed skill in writing an argumentative essay.	The writer generates an argument that productively engages with multiple perspectives on the given issue. The argument's thesis reflects precision in thought and purpose. The argument establishes and employs a thoughtful context for analysis of the	Development of ideas and support for claims deepen understanding. A mostly integrated line of purposeful reasoning and illustration capably conveys the significance of the argument. Qualifications and complications	The response exhibits a productive organizational strategy. The response is mostly unified by a controlling idea or purpose, and a logical sequencing of ideas contributes to the effectiveness of the	The use of language works in service of the argument. Word choice is precise. Sentence structures are clear and varied often. Stylistic and register choices, including voice and tone, are purposeful and productive. While minor errors in grammar, usage, and mechanics may be present,

Score Point	Ideas & Analysis	Development & Support	Organization	Language Use & Conventions
	issue and its perspectives. The analysis addresses implications, complexities and tensions, and/or underlying values and assumptions.	enrich ideas and analysis.	argument. Transitions between and within paragraphs consistently clarify the relationships among ideas.	they do not impede understanding.
Score 4: Responses at this score point demonstrate adequate skill in writing an argumentative essay.	The writer generates an argument that engages with multiple perspectives on the given issue. The argument's thesis reflects clarity in thought and purpose. The argument establishes and employs a relevant context for analysis of the issue and its perspectives. The analysis recognizes implications, complexities and tensions, and/or underlying values and assumptions	Development of ideas and support for claims clarify meaning and purpose. Lines of clear reasoning and illustration adequately convey the significance of the argument. Qualifications and complications extend ideas and analysis.	The response exhibits a clear organizational strategy. The overall shape of the response reflects an emergent controlling idea or purpose. Ideas are logically grouped and sequenced. Transitions between and within paragraphs clarify the relationships among ideas.	The use of language conveys the argument with clarity. Word choice is adequate and sometimes precise. Sentence structures are clear and demonstrate some variety. Stylistic and register choices, including voice and tone, are appropriate for the rhetorical purpose. While errors in grammar, usage, and mechanics are present, they rarely impede understanding.
Score 3: Responses at this score point demonstrate some developing skill in writing an argumentative essay.	The writer generates an argument that responds to multiple perspectives on the given issue. The argument's thesis reflects some clarity in thought and purpose. The	Development of ideas and support for claims are mostly relevant but are overly general or simplistic. Reasoning and illustration largely clarify the argument	The response exhibits a basic organizational structure. The response largely coheres, with most ideas logically grouped. Transitions	The use of language is basic and only somewhat clear. Word choice is general and occasionally imprecise. Sentence structures are usually clear but show little variety.

Score Point	Ideas & Analysis	Development & Support	Organization	Language Use & Conventions
	argument establishes a limited or tangential context for analysis of the issue and its perspectives. Analysis is simplistic or somewhat unclear.	but may be somewhat repetitious or imprecise.	between and within paragraphs sometimes clarify the relationships among ideas.	Stylistic and register choices, including voice and tone, are not always appropriate for the rhetorical purpose. Distracting errors in grammar, usage, and mechanics may be present, but they generally do not impede understanding.
Score 2: Responses at this score point demonstrate weak or inconsistent skill in writing an argumentative essay	The writer generates an argument that weakly responds to multiple perspectives on the given issue. The argument's thesis, if evident, reflects little clarity in thought and purpose. Attempts at analysis are incomplete, largely irrelevant, or consist primarily of restatement of the issue and its perspectives.	Development of ideas and support for claims are weak, confused, or disjointed. Reasoning and illustration are inadequate, illogical, or circular, and fail to fully clarify the argument.	The response exhibits a rudimentary organizational structure. Grouping of ideas is inconsistent and often unclear. Transitions between and within paragraphs are misleading or poorly formed.	The use of language is inconsistent and often unclear. Word choice is rudimentary and frequently imprecise. Sentence structures are sometimes unclear. Stylistic and register choices, including voice and tone, are inconsistent and are not always appropriate for the rhetorical purpose. Distracting errors in grammar, usage, and mechanics are present, and they sometimes impede understanding.
Score 1: Responses at this score	The writer fails to generate an argument that	Ideas lack development, and claims	The response does not exhibit an	The use of language fails to demonstrate skill

Score Point	Ideas & Analysis	Development & Support	Organization	Language Use & Conventions
point demonstrate little or no skill in writing an argumentative essay.	responds intelligibly to the task. The writer's intentions are difficult to discern. Attempts at analysis are unclear or irrelevant.	lack support. Reasoning and illustration are unclear, incoherent, or largely absent.	organizational structure. There is little grouping of ideas. When present, transitional devices fail to connect ideas.	in responding to the task. Word choice is imprecise and often difficult to comprehend. Sentence structures are often unclear. Stylistic and register choices are difficult to identify. Errors in grammar, usage, and mechanics are pervasive and often impede understanding.

Managing Rater Quality

Training and qualification provide initial quality assurance for all raters, but quality monitoring activities continue throughout the performance scoring process. ACT employs several quality assurance processes that establish and maintain consistent calibration and ensure that every response—those scored on the first day through those scored on the last—is given the most appropriate score. ACT's standard quality assurance practices include the following:

- **Reliability Scoring:** Every ACT writing response is reviewed and scored by at least two independent, qualified raters. In cases where scores are nonadjacent, a response is automatically rerouted for a third review by a scoring supervisor or director, and the discrepancy is appropriately resolved. Because of these rigorous training and qualification requirements, nonadjacency rates routinely amount to less than 4% of the overall response population.

- **Validity:** Validity responses are selected and prescored by scoring supervisors and directors and then inserted as part of the workflow. Rater accuracy is measured by rate of agreement with validity responses. A rater whose performance falls below established quality thresholds is excluded from scoring and is subject to retraining activities, including receiving supervisor feedback and taking calibration tests. Raters who fail to demonstrate improved accuracy may be released from the project and their work reset and rescored.

- **Backreading:** The backreading process enables scoring supervisors and directors to review raters' work and provide effective, tailored feedback based on specific scoring

examples. The backreading process also allows for new scores to be applied where necessary. This is an important part of the quality assurance process, and all raters are subject to daily backreading.

- **Calibration:** General and targeted calibration exercises are administered regularly throughout the performance scoring process to maintain rater accuracy and address any emergent scoring trends. Calibration sets are compiled by scoring supervisors and directors to address specific scoring trends or create a retraining exercise for targeted individual raters.

- **Quality Reporting:** ACT utilizes a suite of dynamic, on-demand quality reports to monitor scoring quality and to quickly identify and diagnose scoring issues at the group or individual rater level. On an ongoing basis, scoring supervisors and directors review data showing inter-rater reliability, validity agreement, frequency distribution, scoring rate, backreading agreement, and other important quality metrics. Table 5.2 provides a sample of some of the available reports.

CRASE+ Scoring

For writing tests completed by computer, ACT may replace one (human) reader with CRASE+, ACT's automated essay scoring engine.

The CRASE+ scoring models for ACT Writing were created using around 9,000 reader-scored essays across a variety of prompts. The models produced scores that agreed with human readers at rates that matched or exceeded the scores produced by independent human readers. Details about the training and validation process, including information about model performance across subgroups, can be found in the document *CRASE+ for ACT Writing Technical Report*, available *here*.

Table 5.2. Sample of Quality Reports

Report Name	Description
Daily/Cumulative Inter-Rater Reliability Summary	Group-level summary of both daily and cumulative inter-rater reliability statistics for each day of the scoring project
Frequency Distribution Report	Task-level summary of score-point distribution percentages on both a daily and a cumulative basis
Daily/Cumulative Validity Summary	Summary of agreement for validity reads of a given task on both a daily and a cumulative basis
Completion Report	Breakdown of the number of responses scored and the number of responses in each stage of scoring (first score, second score, resolution)
Performance Scoring Quality Management Report	Summary of task-level validity and inter-rater reliability on a daily and cumulative basis: This report also shows the number of resolutions required and completed, as well as task-level frequency distribution.

5.2.3 Composite, STEM, and ELA Scores

The ACT Composite score represents a student's overall performance on all the multiple-choice test sections. It is the average of the scale scores for English, mathematics, reading, and science rounded to the nearest whole number (decimals 0.5 or greater rounded up). The STEM score represents a student's overall performance on the science and mathematics tests. It is the rounded average of the mathematics and science scale scores. The ELA score represents a student's overall performance on the English, reading, and writing tests. It is the rounded average of the English, reading, and 1–36 writing scale scores. Only students who take the writing test along with the ACT test receive an ELA score. To calculate ELA scores, ACT converts the sum of the writing domain scores to a 1–36 scale. Procedures for obtaining the 1–36 writing scale scores are described in *Chapter 6*. The Composite, STEM, and ELA scores all range from 1 to 36. By virtue of equating, each of these scores is comparable for students who are administered different test forms.

5.2.4 ACT Superscores

The ACT Superscore is the average of the four best test section scores across ACT test attempts. Superscores were first provided to students during the 2020–2021 academic year, and they count as official ACT scores for reporting to colleges and universities. Research on the validity of ACT Superscores is provided in *Chapter 7* of the ACT Technical Manual. To be eligible for an ACT Superscore, a student must complete the full ACT multiple-choice test (English, math, reading, and science) on a single testing occasion. Once a student has taken the ACT multiple times, the highest score in each section is identified, and the four scores are averaged and rounded to the nearest whole number. The same basic process is also carried out to calculate ACT Superscores for STEM and ELA.

5.2.5 Interpretation of the ACT Test Scores

The ACT score reports present additional information to help students and educators interpret scores. This includes standard errors of measurement (SEMs), the ACT College Readiness Benchmarks, and the national and state ranks of the scores.

SEM and Score Ranges

The score report contains information about the measurement precision of the test section, Composite, STEM, and ELA scores. The SEM reflects imprecision in test scores related to the fact that students would not necessarily earn the same scores if they took the ACT repeatedly. The SEMs are about 1 point for the writing and the Composite scores and about 2 points for the test section, STEM, and ELA scores. Students' scores are reported with score ranges that are graphically represented by shaded areas around their scores. Detailed information about measurement precision is given in *Chapter 6*.

ACT College Readiness Benchmarks

The ACT College Readiness Benchmarks are scores that represent the level of achievement associated with at least a 50% chance of earning a B or higher or about a 75% chance of earning a C or higher in specific first-year college courses in the corresponding subject area. A Benchmark is available for each multiple-choice section and the STEM and ELA scores. Students' readiness for first-year college courses corresponding to each multiple-choice test and to STEM and ELA scores can be assessed by comparing students' scores with these Benchmarks. The STEM Benchmark is the minimum STEM score associated with success in first-year college courses in STEM majors, and the ELA Benchmark is the minimum ELA score associated with success in first-year college ELA courses.

Additional resources are available to facilitate interpretating ACT scores. The ACT College and Career Readiness Standards are sets of statements intended to help students, parents, and educators understand the meaning of test scores. These Standards relate test scores to the types of skills needed for success in high school and beyond. They serve as a direct link between what students have learned and what they are ready to do next. To gain insights into the ACT test scores and the Standards, see Sections *5.5* and *5.6* in this chapter for more details about the ACT's College and Career Readiness Standards and ACT College Readiness Benchmarks .

Score Norms

The national (U.S.) and state ranks can help students understand how their scores compare to those of other students in the nation and in their states. A rank indicates the percentage of tested students whose scores are the same as or lower than a given student's score. ACT U.S. and state ranks are based upon the most recent scores of high school seniors who graduated during the previous three years and took the ACT in 10th, 11th, or 12th grade. The most recent U.S. ranks are available at *http://www.act.org/content/act/en/products-and-services/the-act/scores/national-ranks.html*. Because these ranks include scores from students who tested in 10th, 11th, or 12th grade, these ranks are not intended to represent the performance of 12th-grade students nationwide.

An examinee's standing on different tests should be compared using norms rather than scale scores. The scale scores were not constructed to ensure that, for example, a 16 on an English test is comparable to a 16 on a mathematics, reading, or science test. In contrast, the examinee ranks on different tests indicate standings relative to the same comparison group (i.e., the norm group). The ranks can be used for making relative comparisons among examinee performance levels on different subjects.

5.2.6 Summary Statistics, Effective Weights, and Correlations

Operational test data from the test forms administered to Wisconsin state testing in the 2024–2025 academic year were analyzed to obtain descriptive statistics reported in this chapter. This part presents the summary statistics and correlations among the test section scores and the Composite and ELA scores. Effective weights are also reported for each component of the Composite and ELA scores.

Score Distribution Summary Statistics

The summary statistics of the ACT test score distributions for the standard forms are presented in Table 5.3. Table 5.4 provides corresponding statistics for the accommodated form administered in spring 2025 in the state administrations of the ACT.

Table 5.3. Summary Statistics of the ACT Test Score Distributions for the Standard Forms Administered in Wisconsin in Spring 2025

Program	Statistic	English	Math	Reading	Science	Composite	Writing	ELA	STEM
	Mean	18.55	19.37	19.71	19.98	19.53	6.66	18.99	19.93
Standard	SD	6.48	5.68	6.45	5.52	5.50	1.83	5.69	5.31
Form	Skewness	0.49	0.75	0.41	0.43	0.56	-0.41	0.17	0.66
	Kurtosis	-0.27	0.09	-0.54	0.30	-0.25	-0.00	-0.44	0.11

Note: Information in this table can also be found in Appendix *Table 3*.

Table 5.4. Summary Statistics of the ACT Test Score Distributions for the Accommodated Forms Administered in Wisconsin in Spring 2025

Program	Statistic	English	Math	Reading	Science	Composite	Writing	ELA	STEM
	Mean	13.31	15.26	14.94	15.88	14.97	4.47	13.06	15.82
Accommodated	SD	4.75	3.73	5.78	4.70	4.20	1.84	5.07	3.89
Form	Skewness	1.71	2.01	1.45	1.22	1.93	0.39	1.24	1.87
	Kurtosis	3.54	6.33	2.14	2.43	4.23	-0.44	1.56	4.80

Note: Information in this table can also be found in Appendix *Table 3*.

Effective Weights

The Composite, STEM, and ELA scores are rounded averages of test section scores. Specifically, the English, mathematics, reading, and science test scale scores are weighted equally to calculate the Composite score; the mathematics and science scale scores are weighted equally to calculate the STEM score; and the English, reading, and writing scale scores are weighted equally to calculate the ELA score. Calculating scores this way makes the weights used in the calculation ¼ for ACT Composite, ½ for STEM, and ⅓ for ELA scores (often referred to as "nominal" weights).

There are other ways to determine the contributions of test scores to a combined score. Effective weights, for example, are defined as the proportion of the variability of the combined score that can be attributed to a particular test score (Wang & Stanley, 1970). Score covariances are calculated and combined to obtain effective weights. Specifically, the effective weight for a test is calculated by summing the values in the appropriate row of the covariance matrix and dividing the resulting value by the sum of all covariances among the tests using the formula

$$(effective\ weight)_x = \frac{\sum_y cov_{xy}}{\sum_x \sum_y cov_{xy}}$$

where cov_{xy} is the covariance of test scores corresponding to row x and column y in the covariance matrix.

For example, to obtain effective weights for the four multiple-choice tests used to calculate the Composite score, ACT computed scale score covariances from one test form administered in the 2024–2025 academic year (see Table 5.5). The effective weight for the English test was computed by adding the four numbers in the first row (42.57, 28.92, 35.20, and 28.79). This number was then divided by the sum of all covariances for all four multiple-choice tests (i.e., the variance of the Composite score), which resulted in an effective weight of 0.27 (after rounding). The effective weights for the mathematics, reading, and science tests were obtained in a similar fashion.

Table 5.6 shows the ranges of effective weights for the Composite and ELA scores based on the Wisconsin students taking test forms administered in the 2024–2025 academic year. For these scores, the effective weights were fairly stable across forms. For the Composite score, the effective weights for the English and reading tests were the largest. They were relatively high because the English and reading tests had the largest score variances and because their covariances with the other measures tended to be the highest. The larger score variances and covariances for the English test also contributed to higher effective weights for English in the ELA score.

Table 5.5. Scale Score Covariances for Multiple-Choice Tests from the Primary Form Administered in Wisconsin in Spring 2025

Test	English	Mathematics	Reading	Science
English	42.57	28.92	35.20	28.79
Mathematics	28.92	33.89	26.83	25.78
Reading	35.20	26.83	40.71	28.32
Science	28.79	25.78	28.32	30.99

Note: Information in this table can also be found in *Appendix Table 4*.

Table 5.6. Effective Weights of the ACT Tests from the Primary Form Administered in Wisconsin in Spring 2025

Test	Composite	ELA
English	0.27	0.35
Mathematics	0.23	--
Reading	0.26	0.33
Science	0.23	--
Writing	--	0.32

Note: Information in this table can also be found in *Appendix Table 5.*

Correlations

Table 5.7 shows the correlations among the ACT test scores based on operational data from the test forms administered in the 2024–2025 academic year in the Wisconsin state-sponsored administration of the ACT. The correlations between the writing scores and other scale scores were relatively low, which was attributable to the smaller range and lower reliability of the writing test scores than the other scores. Score reliability of the ACT tests is presented in *Chapter 6*.

Table 5.7. Correlations Among the ACT Test Scores Administered in Wisconsin in Spring 2025

Score	English	Mathematics	Reading	Science	Composite	Writing	ELA
English	1.00	0.75	0.82	0.77	0.93	0.56	0.91
Mathematics		1.00	0.70	0.79	0.88	0.49	0.74
Reading			1.00	0.77	0.91	0.53	0.89
Science				1.00	0.91	0.51	0.78
Composite					1.00	0.57	0.92
Writing						1.00	0.81
ELA							1.00

Note: Information in this table can also be found in *Appendix Table 6*.

5.3 Detailed Performance Description

As shown in Figures 5.3 and 5.4, ACT score reports include detailed results that describe students' performance on finer-grained skills and domains within each test section. This includes reporting category scores and ACT Readiness ranges for each multiple-choice test as well as domain scores for the ACT writing test.

Figure 5.3. Detailed Results on a Sample Interactive Score Report on MyACT

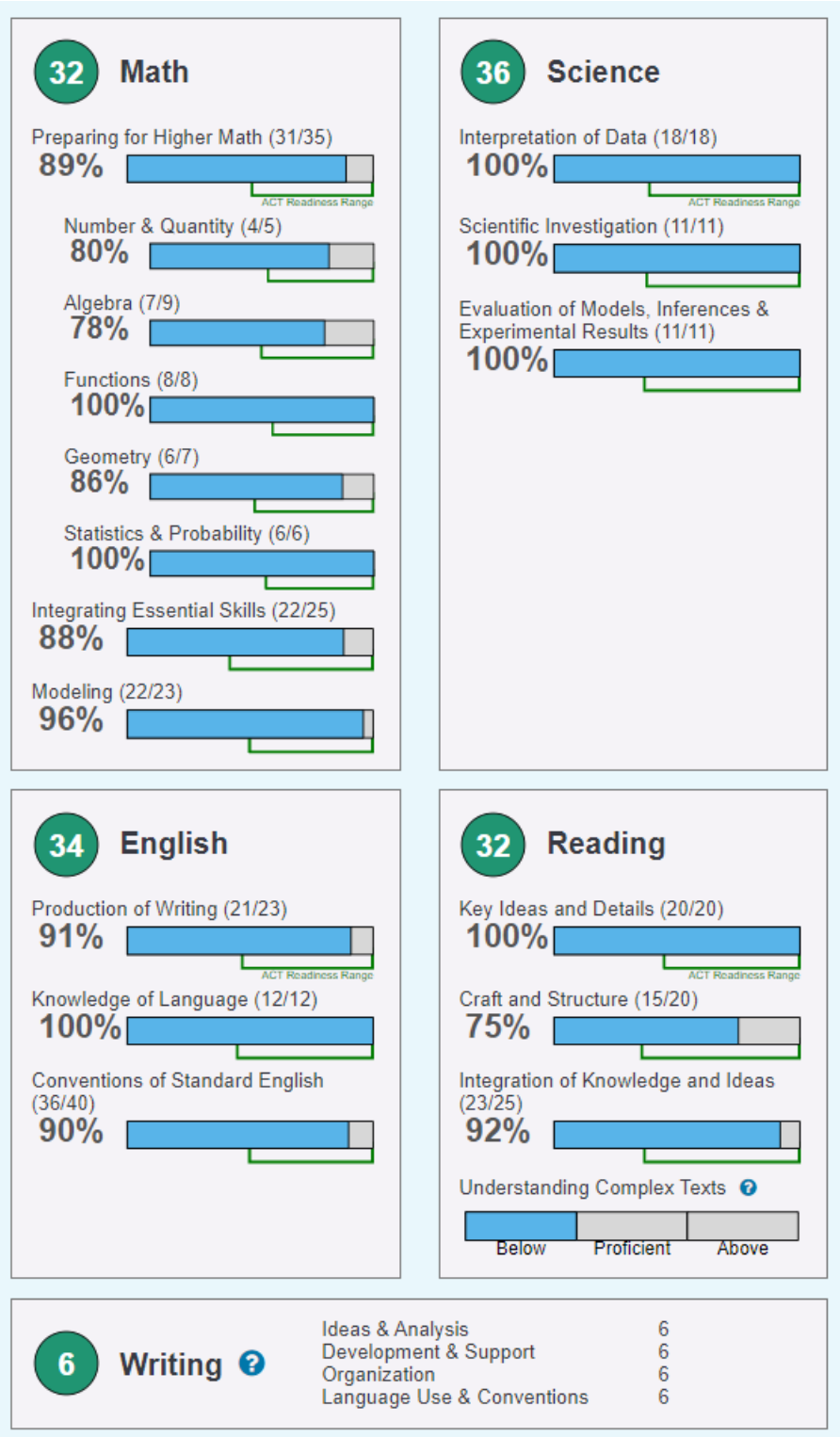
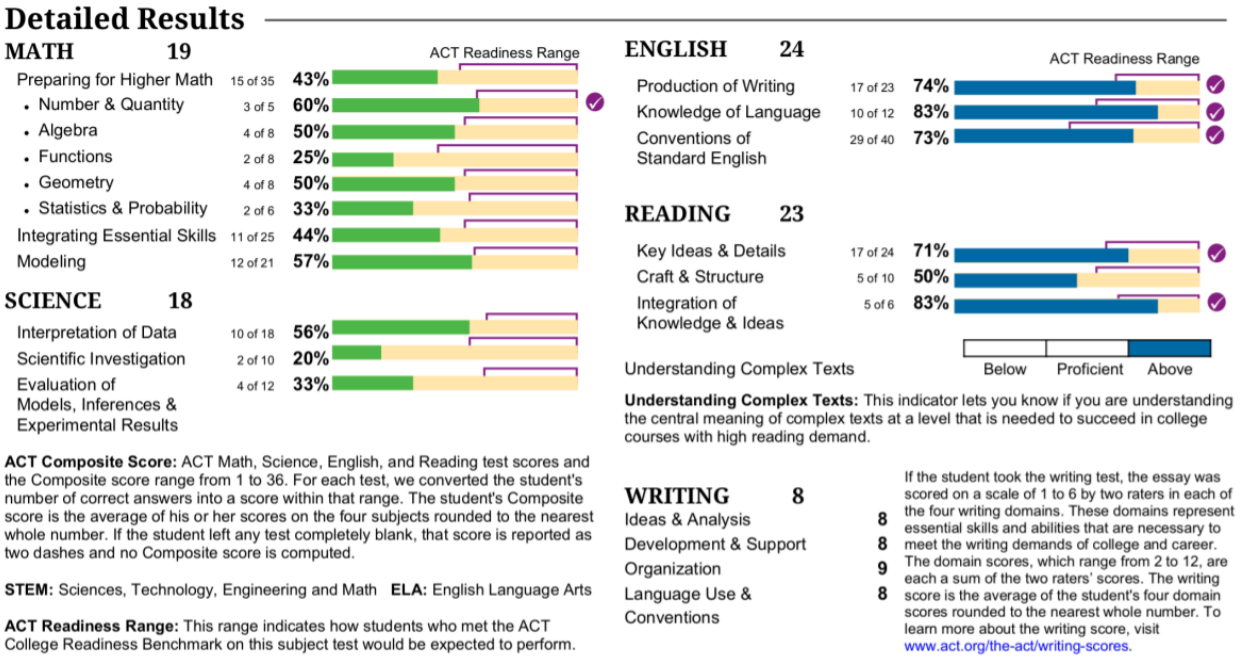


Figure 5.4. Detailed Results on a Sample ACT High School Score Report



5.3.1 Reporting Categories and ACT Readiness Ranges

ACT reporting categories are aligned with the ACT College and Career Readiness Standards (see *Section 5.5*) and other standards that target college and career readiness. Items that measure similar skills are grouped together to provide students with more detailed information about their test performance within each subject. There are three reporting categories each for English, reading, and science and eight for mathematics. These reporting categories make it easier for students, parents, and educators to gain insight into students' performance by highlighting students' relative strengths and areas for improvement in each subject. Beginning in fall 2016, reporting category scores replaced the subscores that were reported previously.

For each reporting category, the total number of points possible, the total number of points a student obtained, and the percentage of points achieved are shown. In addition, for each reporting category, there is an ACT Readiness Range indicating the expected percentage correct scores for students who scored at or above the ACT College Readiness Benchmark for that specific subject. Note that the number of items for a particular reporting category can vary across different test forms. The Readiness Ranges vary accordingly, and they also account for differences in reporting category item difficulty across forms following the procedure described in *Chapter 6*.

Information about the development and blueprints of ACT reporting categories is in *Chapter 3*. Details about interpreting ACT reporting categories and ACT Readiness Ranges are in the *ACT Reporting Category Interpretation Guide* (Powers, Li, Suh, & Harris, 2016).

5.3.2 Writing Domain Scores

In addition to the overall writing test score, scores are also reported for four domains: Ideas & Analysis, Development & Support, Organization, and Language Use & Conventions. These domains reflect essential skills and abilities that are required for college and career success. Each essay is scored on a scale of 1 to 6 by two raters (one of which may be CRASE+) on each of the four domains. If the scores from the two raters differ by more than 1 point on any of the domains, a third rater evaluates the essay to resolve the discrepancy. A domain score, ranging from 2 to 12, is the sum of the two raters' scores. Detailed descriptions of the writing domains and the analytic scoring rubric used to score the writing test are in *Chapter 3*.

Table 5.8 presents the summary statistics of writing domain scores and the overall writing scores based on ACT writing test forms administered in the 2024–2025 academic year in the Wisconsin state-sponsored administration of the ACT. Table 5.9 presents the correlations among these scores for ACT Wisconsin testers.

Table 5.8. Summary Statistics of the ACT Writing and Writing Domain Score Distributions for the Wisconsin Spring 2025 Administration

Statistic	Ideas & Analysis	Development & Support	Organization	Language Use & Conventions	Writing Score
N	58,220	58,220	58,220	58,220	58,220
Mean	6.63	6.24	6.52	6.88	6.66
SD	1.91	1.89	1.87	1.74	1.83
Skewness	-0.44	-0.25	-0.48	-0.43	-0.41
Kurtosis	-0.03	-0.36	-0.08	0.30	-0.00

Note: Information in this table can also be found in *Appendix Table 7*.

Table 5.9. Correlations Among the ACT Writing and Writing Domain Scores for the Wisconsin Spring 2025 Administration

Score	Ideas & Analysis	Development & Support	Organization	Language Use & Conventions	Writing Score
Ideas & Analysis	1.00	0.95	0.98	0.95	0.99
Development & Support		1.00	0.95	0.91	0.95
Organization			1.00	0.94	0.98
Language Use & Conventions				1.00	0.97
Writing Score					1.00

Note: Information in this table can also be found in *Appendix Table 8*.

5.3.3 Understanding Complex Texts Indicator

The Understanding Complex Texts (UCT) indicator is reported to show whether students understand the central meaning of complex texts at the level needed to succeed in college courses with higher reading demands. This indicator is based on scores on a subset of items on

the reading test. These items measure students' global comprehension of the passages instead of sentence- or word-level understanding. Students' overall performance on these items is classified into three levels: Below Proficient, Proficient, and Above Proficient.

The performance levels were first established through a special study that linked students' scores on UCT items to their college course grades (Allen, Bolender, Fang, Li, & Thompson, 2016). This special study examined the UCT scores and course grades of 263,265 students from 439 postsecondary institutions. To obtain UCT scores for the study, content experts classified the UCT test items retroactively for each form so that students' number correct UCT scores could be calculated. The number of items that contributed to the UCT score varied across forms. The number correct UCT scores were then equated across forms to obtain an interim score scale ranging from 0 to 16.

As expected, results of the special study indicated that the UCT scores were more predictive of success in college courses that have higher demand for understanding complex texts. Hierarchical logistic regression was used to model the relationship between UCT scores and students' chances of earning a B or higher grade in seven types of courses (American History*, Literature, other history*, other natural science, Physics without Calculus, Sociology, and Zoology*). Three of the seven course types (marked with *) were also used to develop the ACT College Readiness Benchmark for reading. The UCT score associated with a 50% chance of earning a B or higher grade was identified for each course and institution. These results were aggregated over a weighted sample of institutions to identify the Proficient cut score of 9 out of 16. The Proficient cut score is also associated with a 78% chance of earning a C or higher and a 22% chance of earning an A.

The Above Proficient cut score of 13 out of 16 was identified in a similar way. This score is associated with a 67% chance of earning a B or higher grade at a typical institution. The Above Proficient cut score is also associated with an 85% chance of earning a C or higher grade and a 37% chance of earning an A. The Above Proficient cut score is about 2 SEMs above the Proficient cut score. For additional information on the development of the UCT cut scores, see the full report *Relating the ACT Indicator Understanding Complex Texts to College Course Grades* by Allen et al. (2016).

5.4 Progress Toward the ACT WorkKeys National Career Readiness Certificate Indicator

The Progress Toward the ACT WorkKeys NCRC indicator is based on students' ACT Composite scores. This indicator provides an estimate of students' most likely performance on the ACT® WorkKeys® National Career Readiness Certificate® (NCRC®), which is an assessment-based credential that certifies foundational work skills important for job success across industries and occupations. The WorkKeys NCRC is based on the results of three assessments: ACT® WorkKeys® Applied Math, ACT® WorkKeys® Workplace Documents, and ACT® WorkKeys® Graphic Literacy. Scores on these assessments determine whether an individual earns a Bronze, Silver, Gold, or Platinum certificate or does not earn a certificate. The WorkKeys NCRC gives individuals evidence that they possess the skills that employers deem

essential to workplace success. Find more information about the WorkKeys NCRC at *http://www.act.org/content/act/en/products-and-services/workkeys-for-educators/ncrc.html*.

Data from nearly 79,000 11th and 12th graders who took the ACT and all three WorkKeys NCRC assessments during the 2017–2018 academic year were used to establish a link between ACT Composite scores and the WorkKeys NCRC levels (Radunzel & Fang, 2018). Logistic regression was used to identify the ACT Composite score that corresponded to at least a 50% chance of obtaining each WorkKeys NCRC level. This method of determining cut scores was similar to the approach used to establish the ACT College Readiness Benchmarks (Allen, 2013). The study showed that the ACT Composite scores corresponding to the Bronze, Silver, Gold, and Platinum certificates were 13, 17, 22, and 27, respectively.

Based on the ACT Composite cut scores obtained for each WorkKeys NCRC level from the linking study, the Progress Toward the ACT WorkKeys NCRC indicator classifies students into one of five levels[2]: unlikely to earn a WorkKeys NCRC (below 13), most likely to earn a Bronze NCRC (13–16), most likely to earn a Silver NCRC (17–21), most likely to earn a Gold NCRC (22–26), and most likely to earn a Platinum NCRC (27–36).

Note that this indicator is not a substitute for an actual WorkKeys NCRC level obtained by taking WorkKeys Assessments. Given the probability-based nature of the indicator and the corresponding uncertainty in the predictions, actual performance on the WorkKeys NCRC can differ from the predicted performance based on the ACT test. Moreover, there are differences in the constructs measured and the content assessed between the two assessments. That said, the Progress Toward the WorkKeys NCRC indicator provides students who take the ACT with some information about their level of career readiness based on academic achievement test results.

5.5 ACT College and Career Readiness Standards

The purpose of this part is to provide background on the ACT College and Career Readiness Standards—for example, their purpose, how they are developed and maintained, and how to interpret them. These Standards are empirically derived descriptions of the essential skills and knowledge students need to become ready for college and career. Parents, teachers, counselors, and students use the Standards to:

- communicate widely shared learning goals and expectations
- relate test scores to the skills needed in high school and beyond

[2] These cut scores and an indicator for the Platinum WorkKeys NCRC were first included on ACT score reports in fall 2018. Note that the ACT cut scores for the Gold and Platinum WorkKeys NCRC progress indicators are lower than those reported prior to fall 2018 (see Allen, LeFebvre, & Mattern, 2016, for information on prior cut scores). As a result of these changes, a larger percentage of students will be identified as most likely to obtain the Gold or Platinum WorkKeys NCRC.

- understand the increasing complexity of skills needed across the score ranges in English, mathematics, reading, science, and writing

The ACT College Readiness Benchmarks are the minimum ACT scores required for students to have a reasonable chance of success in credit-bearing college courses—English Composition I, social sciences courses, College Algebra, or Biology (see *Section 5.6*).

5.5.1 Description of the ACT College and Career Readiness Standards

In 1997, ACT began an effort to make the ACT test results more informative and useful. This effort yielded the ACT College and Career Readiness Standards, which are statements that describe what students who score in various score ranges on the tests are likely to know and be able to do. For example, students who score in the 16–19 range on the ACT English test typically are able to "determine the most logical place for a sentence in a paragraph," whereas students who score in the 28–32 score range are able to "determine the most logical place for a sentence in a fairly complex paragraph." These Standards reflect a progression of skills in each of the five test sections: English, mathematics, reading, science, and writing. ACT organized the Standards by strands—related areas of knowledge and skills within each test—to be easier for teachers and curriculum specialists to use. The complete Standards are posted on ACT's website: *www.act.org/content/act/en/college-and-career-readiness/standards.html*.

The Standards are provided for six score ranges along the 1–36 score scale for the ACT test. Students who score in the 1–12 range are most likely beginning to develop the knowledge and skills described in the 13–15 score range. The Standards are cumulative, which means that if students score, for example, in the 20–23 range on the English test, they are likely to be able to demonstrate most or all of the knowledge and skills described in the preceding score ranges.

ACT developed the Standards for the writing test in 2005 and updated them with enhancements in 2015. The writing test Standards are provided for five score ranges in four writing domains based on ACT writing test scores (the sum of two raters' scores according to the 6-point analytic scoring rubric for the ACT writing test). Scores below 3 in any domain on the writing test do not permit useful generalizations about students' writing abilities. That is, students scoring in this range provide little evidence of writing skills relevant to that domain.

5.5.2 Determining the Score Ranges for the ACT College and Career Readiness Standards

When ACT began work on the College and Career Readiness Standards in 1997, the first step was to determine the number of score ranges and the width of each score range. To do this, ACT staff reviewed the ACT normative data in the context of how the test scores are used—for example, the use of the ACT scores in college admissions and course-placement decisions.

In reviewing the normative data, ACT staff analyzed the distribution of student scores across the ACT score scale (1–36) and reevaluated course placement research that ACT had conducted over the previous 40 years. In the past, ACT's Course Placement Service provided colleges and universities with cutoff scores used for placement into appropriate entry-level college courses.

Cutoff scores based on admissions and course-placement criteria were used to help define the score ranges for the four multiple-choice test sections.

After analyzing all the data and reviewing different possible score ranges, ACT staff concluded that the score ranges 1–12, 13–15, 16–19, 20–23, 24–27, 28–32, and 33–36 would best distinguish students' levels of achievement so as to assist teachers, administrators, and others to relate the ACT multiple-choice test scores to students' skills and knowledge.

5.5.3 Developing the ACT College and Career Readiness Standards

After reviewing the normative data, college admissions criteria, and ACT scores associated with success in postsecondary courses obtained through ACT's Course Placement Service (a service no longer offered), subject matter experts wrote the ACT College and Career Readiness Standards based on their analysis of the skills and knowledge students need in order to respond successfully to test items that were answered correctly by 80% or more of the examinees who scored within each score range. Content specialists analyzed test items taken from dozens of test forms. The 80% criterion was chosen because it offers those who use the Standards a high degree of confidence that students scoring within a given score range will most likely be able to demonstrate the skills and knowledge described in that range.

Process

Four ACT content teams were identified, one for each of the multiple-choice tests (English, mathematics, reading, and science). Each content team was provided with numerous test forms and data showing the percentages of students in each score range who answered each test item correctly (i.e., item difficulty by student group scoring within the score range). For example, the mathematics content team reviewed 10 forms of the ACT mathematics test. There are 60 items in each ACT mathematics test form, so 600 ACT mathematics items were reviewed in all.

An illustrative table displaying the information provided to the mathematics content team for one ACT mathematics test form is shown in Table 5.10. The shaded areas in this table show the items that met the 0.80-or-above item difficulty criterion for each of the score ranges. As illustrated in the table, a cumulative effect can be noted. That is, the items that were correctly answered by 80% of the students in the 16–19 score range also appear in the 20–23 score range, and so on. By using this information, the content teams were able to isolate and review the items by score ranges across test forms. Table 5.11 reports the total number of test items reviewed for each content area.

Table 5.10. Illustrative Listing of Mathematics Item Difficulties by Score Range

	Score Range					
Item no.	13–15	16–19	20–23	24–27	28–32	33–36
1	.62	.89	.98	.99	1.00	1.00
2		.87	.98	.99	.99	1.00
6	.60	.86	.94	.97	.99	.99
7	.65	.92	.98	.99	.99	1.00
20		.84	.94	.97	.98	.99
27		.85	.97	.99	.99	.99
4			.92	.97	.99	1.00
5			.94	.97	.99	.99
⋮			⋮	⋮	⋮	⋮
8			.82	.95	.98	.99
9			.80	.89	.96	.99
21			.82	.92	.97	.99
13				.90	.97	.99
15				.90	.97	.99
39				.85	.96	.99
44				.84	.96	.99
25					.95	.99
28					.97	1.00
⋮					⋮	⋮
35					.86	.96
47					.86	.97
32						.95
46						.90
49						.95
⋮						⋮

Table 5.11. Number of ACT Items Reviewed During 1997 National Review

Content Area	Number of Items for Each Test
English	75
Mathematics	60
Reading	40
Science	40
Number of items per form	215
Total number of test forms reviewed	10
Total number of items reviewed	2,150

These procedures allowed the content teams to conceptualize what each ACT test section measures. Specifically, each content team followed the same process as they reviewed the items in each ACT multiple-choice test:

1. Multiple forms of each test were distributed.

2. The skills and knowledge necessary to answer the test items in the lowest score range were identified.

3. The additional skills and knowledge necessary to answer the test items in the next (higher) score range were identified. This step was repeated for all remaining score ranges.

4. All the lists of statements identified by each content specialist were merged into a composite list. The composite list was distributed to a broader group of content specialists.

5. The composite list was reviewed by each content specialist, and ways to generalize and consolidate the various skills and knowledge were identified.

6. The content specialists met as a group to discuss the individual, consolidated lists and prepared a master list of skills and knowledge, organized by score ranges.

7. The master list was used to review at least three additional test forms, and adjustments and refinements were made as needed.

8. The adjustments were reviewed by the content specialists, and revisions were made.

9. The list of skills and knowledge was used to review additional test forms. The purpose of this review was to determine whether the Standards adequately and accurately described the skills and knowledge measured by the items specific to each score range.

10. The ACT College and Career Readiness Standards were further refined, as needed, and finalized.

Conducting an Independent Review of the ACT College and Career Readiness Standards

As a means of gathering content validity evidence, ACT invited nationally recognized scholars in English, mathematics, reading, science, and education departments from high schools and universities to review the ACT College and Career Readiness Standards. These teachers and researchers were asked to provide ACT with independent, authoritative reviews of the Standards. The selection process sought and achieved a diverse representation by gender, ethnic background, and geographic location. Each participant had extensive and current knowledge of his or her field, and many had acquired national recognition for their professional accomplishments.

The reviewers were asked to evaluate whether the Standards (a) accurately reflected the skills and knowledge needed to correctly respond to test items (in specific score ranges) on the ACT and (b) represented a continuum of increasingly sophisticated skills and knowledge across the score ranges. Each national content area team consisted of three college faculty members currently teaching courses on curriculum and instruction (in schools of education) and three classroom teachers, one each from eighth, 10th, and 12th grades. The reviewers were provided

with the complete set of Standards and a sample of test items falling within each of the score ranges for each test.

The samples of items to be reviewed by the consultants were randomly selected for each score range in all four multiple-choice tests. ACT believed that a random selection of items would ensure a more objective outcome than would preselected items. Ultimately, 17 items for each score range were selected. Before identifying the number of items that would comprise each set of items in each score range, it was first necessary to determine the target criterion for the level of agreement among the consultants. ACT decided upon a target criterion of 70%. It was deemed most desirable for the percentage of matches to be estimated with an accuracy of plus or minus 5%. That is, the standard error of the estimated percent of matches to the Standards should be no greater than 5%. To estimate a percentage around 70% with that level of accuracy, 85 observations were needed. Since there were five score ranges, the number of items per score range to be reviewed was 17 (85 ÷ 5 = 17).

The consultants had two weeks to review the ACT College and Career Readiness Standards. Each reviewer received a packet of materials that contained the Standards, sets of randomly selected items (17 per score range), introductory materials about the Standards, a detailed set of instructions, and two evaluation forms.

The sets of materials submitted for the experts' review were drawn from 13 ACT forms. The consultants were asked to perform two main tasks in their areas of expertise: Task 1—Judge the consistency between the Standards and the corresponding sample items provided for each score range; and Task 2—Judge the degree to which the Standards represent a cumulative progression of increasingly sophisticated skills and knowledge from the lowest score range to the highest score range. The reviewers were asked to rate the items using a 5-point Likert scale that ranged from "strongly agree" to "strongly disagree." They were also asked to suggest revisions to the language of the Standards that would help them better reflect the skills and knowledge measured by the sample items.

ACT collated the consultants' ratings and comments as they were received. The consultants' reviews in all but two cases reached ACT's target criterion, as shown in Table 5.12. That is, 70% or more of the consultants' ratings were "agree" or "strongly agree" when judging whether the Standards adequately described the skills required by the test items and whether the Standards adequately represented the cumulative progression of increasingly sophisticated skills from the lowest to the highest score ranges. The one exception was the ACT reading test, where the degree of agreement was 60%. Each ACT staff content area team met to review all comments made by the national consultants. The teams reviewed all suggestions and adopted several helpful clarifications in the language of the Standards, particularly in the language of the ACT reading test Standards in which the original language failed to meet the target criterion.

Table 5.12. Percentage of Agreement of 1997 National Expert Review

Subject	Task 1	Task 2
English	75%	86%
Mathematics	95%	100%
Reading	60%	100%
Science	70%	80%

5.5.4 The ACT College and Career Readiness Standards for Writing

The score ranges and the ACT College and Career Readiness Standards for the writing test were derived from the ACT writing test scoring rubric. The writing test scoring rubric is a four-domain, 6-point descriptive scale to which writing essays are compared in order to determine their scores (see *Table 5.1 in Section 5.2.2*). Each essay written for the writing test is scored by two trained raters, each of whom gives it a rating from 1 (low) to 6 (high) for each of the four domains. The sum of those two ratings for the domain is a student's writing test domain score (ranging from 2 to 12).

The writing domains assessed by the ACT writing test correspond to key dimensions of effective writing that are taught in high school and college-level composition courses: Ideas & Analysis, Development & Support, Organization, and Language Use & Conventions. These writing domains replace the previous five strands of the ACT College and Career Readiness Standards for Writing, which were derived from a holistic scoring rubric. The design of the enhanced writing test and accompanying Standards reflects the input of several independent consultants, including high school and postsecondary instructors, as well as results from the ACT National Curriculum Survey®.

To determine the score ranges for the writing Standards, ACT staff considered the differences in writing ability evident in essays between levels of the scoring rubric. Based on similarities found among written responses at certain adjacent score points, ACT staff determined that the five score ranges would best distinguish students' levels of writing achievement to assist teachers, administrators, and others to relate ACT test scores to students' skills and knowledge. Writing that receives a score of 2 or lower does not permit useful generalizations about the student's writing abilities in that domain.

5.5.5 Periodic Review of the ACT College and Career Readiness Standards

ACT has conducted periodic internal reviews of its College and Career Readiness Standards. For those reviews, ACT identified three to four new forms of the ACT and then analyzed the data and the corresponding test items specific to each score range. Topics were also compared to data from the most recent ACT National Curriculum Survey (e.g., ACT, 2020). The purposes of these reviews were to ensure that the Standards reflected (a) the most important knowledge and skills for college and career readiness, (b) what was being measured by the items in each score range, and (c) a cumulative progression of increasingly sophisticated skills and knowledge from the lowest score range to the highest score range. Minor refinements intended to update and clarify the language of the Standards resulted from these reviews.

5.5.6 Interpreting and Using the ACT College and Career Readiness Standards

Because new ACT test forms are developed on a regular basis and because no one test form measures all the skills and knowledge included in any particular standard, the ACT College and Career Readiness Standards must be interpreted as knowledge and skills that most students who score within a particular score range are likely to be able to demonstrate. Since there were relatively few test items that were answered correctly by 80% or more of the students who scored in the lower score ranges, the Standards in these ranges should be interpreted with caution.

ACT tests include items measuring areas of knowledge and a large domain of skills that have been judged important for success in high school, college, and beyond. Thus, the Standards should be interpreted in a responsible way that will help students, parents, teachers, and administrators do the following:

- Identify skill areas in which students might benefit from further instruction.
- Monitor student progress and modify instruction to accommodate learners' needs.
- Encourage discussion among principals, curriculum coordinators, and classroom teachers as they evaluate their academic programs.
- Enhance discussions between educators and parents to ensure that students' course selections are appropriate and consistent with their plans after high school.
- Enhance the communication between secondary and postsecondary institutions.
- Identify the knowledge and skills that students entering their first year of postsecondary education should know and be able to do in the academic areas of English language arts, mathematics, and science.
- Assist students as they identify skill areas they need to master to prepare for college-level coursework.

5.6 ACT College Readiness Benchmarks

5.6.1 Description of the ACT College Readiness Benchmarks

The ACT College Readiness Benchmarks are the ACT scores that represent the level of achievement required for students to have a 50% chance of obtaining a B or higher or about a 75% chance of obtaining a C or higher in corresponding credit-bearing first-year college courses at a typical 2-year or 4-year postsecondary institution (Table 5.13). For example, the ACT English Benchmark (18) is the score associated with having a 50% chance of obtaining a B or higher grade in English Composition I.

Table 5.13. ACT College Readiness Benchmarks

College Course(s) or Course Area	ACT Test Score	ACT Benchmark
English Composition I	English	18
College Algebra	Mathematics	22
American History, other history, Psychology, Sociology, Political Science, and Economics	Reading	22
Biology	Science	23
Calculus I, Biology, Chemistry, Physics, and Engineering	STEM	26
English Composition I and social science courses	ELA	20

Three separate studies were conducted to develop the current Benchmarks, and the reports documenting those studies provide more details on the study methodology and samples. The first study developed the ACT Benchmarks in English, reading, mathematics, and science (Allen, 2013). The second study developed the STEM Benchmark (Radunzel, Mattern, Crouse, & Westrick, 2015), and the third study developed the ELA Benchmark (Radunzel, Westrick, Bassiri, & Li, 2017).

Benchmarks were developed for the courses or course combinations listed in Table 5.13. Success in a course was defined as earning a grade of B or higher in the course. Hierarchical logistic regression was used to model the probability of success in a course as a function of ACT test score within each college. The student-level data were weighted to make the sample more representative of all ACT-tested students. For each course within each college, a cutoff score was chosen such that the probability of success (i.e., the probability of earning a B or higher grade in the course) was at least 0.50. This score point most accurately classified the sample into those who would be successful and those who would not (Sawyer, 1989b). The individual cutoff scores per college were weighted to make the sample more representative of all colleges with respect to institution type and selectivity (2-year, 4-year less selective, and 4-year more selective). The Benchmarks (Table 5.13) were determined on by the median cutoff scores across colleges.

5.6.2 Intended Uses of the ACT College Readiness Benchmarks

We recommend that the ACT College Readiness Benchmarks be used for any of three general purposes:

1. Identifying students who are ready for credit-bearing courses (e.g., for course placement) or who need additional academic support (e.g., for early identification for intervention): Because success in college courses depends on more than just the knowledge and skills measured by the ACT test, the best course placement and early identification systems use multiple measures, such as high school GPA, ACT test scores, high school courses taken, and measures of social and emotional learning. The Benchmarks can be used to identify students who have the requisite knowledge and skills targeted by the ACT test. Because performance expectations and grading standards vary across colleges, the Benchmarks represent a standard for the typical postsecondary institution.

2. Serving as a performance standard for K–12 students: The Benchmarks can help states, districts, and schools identify the levels of performance on academic achievement tests that are needed for a student to be ready for college and career. The Benchmarks help articulate college expectations not only to students in high school but also to students in lower grades. Assessments designed for lower grades (e.g., PreACT and PreACT 8/9) can use the ACT test as the anchor of the assessment system and use the Benchmarks as the end target. Some states use the ACT test and the Benchmarks for federal or state accountability reporting.

3. Monitoring educational improvement and achievement gaps over time: Educational stakeholders at all levels (school, district, state, nation) are interested in how their institutions are improving and in the extent that gaps between student groups change over time. The percentage of students meeting the Benchmarks can be used as one of the metrics for monitoring progress and setting goals, and it is most relevant when the ACT test is administered to all students. One advantage of using the Benchmarks for this purpose is that they are indicators of readiness for college coursework and so have relevance to students, educators, and policymakers.

5.6.3 Interpreting ACT Test Scores with Respect to Both the ACT College and Career Readiness Standards and ACT College Readiness Benchmarks

The performance levels on the ACT test necessary for students to be ready to succeed in college-level work are defined by the ACT College Readiness Benchmarks. Meanwhile, the knowledge and skills a student currently has (and areas for improvement) can be identified by examining the student's ACT test scores with respect to the ACT College and Career Readiness Standards. These two empirically derived metrics are designed to help a student translate test scores into a clear indicator of the student's current level of college readiness and to help the student identify key knowledge and skill areas that are needed to improve the likelihood of achieving college success.

Chapter 6
Scaling, Equating, and Technical Characteristics

This chapter discusses the construction of the score scales and the procedures for equating the ACT® tests. The scaling and equating of the multiple-choice tests are described first, followed by the scaling and equating of the ACT writing test scores used for the ELA score calculation. This is followed by a reporting of the psychometric properties of the annual administrations of the ACT and a discussion of comparability between scores from paper and online test administrations.

6.1 Scaling and Equating of the ACT English, Mathematics, Reading, and Science Tests

6.1.1 The Scaling Process

The data used in the scaling process were collected in the fall of 1988 as part of the Academic Skills Study, which provided data to revise the score scale and develop nationally representative norms. Over 100,000 high school students participated in the study. A nationally representative sample of 12th-grade college-bound examinees was used in scaling the ACT. A detailed discussion of the data used for scaling the ACT is given by Sawyer (1989a).

The scaling process for the ACT consisted of three steps. First, weighted raw score distributions for college-bound examinees from the Academic Skills Study were computed. Second, the weighted raw score distributions were smoothed with a four-parameter beta compound binomial model (Lord, 1965; Kolen, 1991; Kolen & Hanson, 1989), and a double arcsine transformation was applied to equalize error variance across the score scale (Kolen, 1988). Finally, the smoothed and arcsine transformed raw score distributions for 12th-grade college-bound examinees were linearly transformed to produce the score scales. These steps are described in greater detail below and by Kolen and Hanson (1989).

In the second step, smoothing of the raw score distributions produced distributions that were easier to work with and that better estimated population distributions. Kolen (1991) and Hanson (1990) showed that smoothing techniques have the potential to improve the estimation of population distributions. Overall, the smoothing process resulted in distributions that appeared smooth without departing much from the unsmoothed distributions. In addition, the first three central moments (mean, variance, and skewness) of the smoothed distributions were identical to those of the original distributions. Values of the fourth central moment of the smoothed distributions (kurtosis) were either identical or very close to those of the original distributions. The double arcsine transformation was applied to the smoothed raw scores to stabilize error variance. This ensured that the conditional standard error of measurement (CSEM) was approximately equal throughout the score scale for 12th-grade college-bound examinees from the Academic Skills Study.

The final step in constructing the score scales was to produce initial scale scores with a specified mean and a specified standard error of measurement (SEM). Methods introduced by Kolen (1988) and described in detail by Kolen and Hanson (1989) were used for this process. After a linear transformation to set the mean score to 18 and the SEM as close to 2 as possible, the initial scale scores were rounded to integers ranging from 1 to 36. Some adjustment of the rounded scale scores was performed to better meet the specified mean and SEM and to avoid gaps in the score scale (i.e., unused scale scores) or to avoid having too many raw scores convert to a single scale score.

In a special study conducted in 1995, the mathematics score scale was reexamined under the condition of allowing calculators (previously calculators had been prohibited on the test). In this study, scores from the mathematics test with calculators were linked to scores from the mathematics test without calculators. It was determined that the score scale created in 1988 would continue to have the same meaning with or without the allowance of calculators on the mathematics test.

6.1.2 Score Scale Characteristics

The scale score range is 1 to 36 for the ACT multiple-choice tests and the Composite, STEM, and ELA scores. The target means of the ACT score scales were 18 for each of the four multiple-choice tests and the Composite for students at the beginning of 12th grade nationwide in 1988 who reported that they planned to attend a two- or four-year college.

Although the score scales for the current ACT tests (administered beginning in October 1989) and the score scale for the original ACT tests (from the ACT's inception in 1959 through all administrations prior to October 1989) are similar, scale scores on these two assessments are not directly comparable due to changes in test content, number of items, test duration, and scaling methodology (e.g., mean score, CSEM, and number of scale points).

For the current ACT, the standard error of measurement was set to be approximately two scale score points for each of the multiple-choice test scores and one scale score point for the Composite. The method described by Kolen (1988) was applied to produce score scales with approximately equal CSEMs along the entire range of scores. If CSEMs were not similar throughout the score scale, CSEMs at different score levels would need to be presented and considered in the interpretation of scores (see AERA, APA, & NCME, 2014, p. 39). Instead, the reported SEM values give reasonably good estimates of the measurement error at all score levels.

The reported scale score for an examinee is only an estimate of that examinee's true scale score. The true score can be interpreted as the average score obtained over countless repeated administrations of the test under identical conditions. If one SEM (approximately two points) was added to and subtracted from each score from repeated administrations, about 68% of the resulting intervals would contain the examinee's true score. This statement assumes a normal distribution for measurement error. The 68% confidence intervals can also be viewed in terms of groups of examinees. Specifically, if one SEM was added to and subtracted from the reported score of each examinee in a group of examinees, the resulting intervals would contain the true

scores for approximately 68% of the examinees. Put another way, about 68% of the examinees would have observed scores that differed from their true scores by less than one SEM. Again, such statements assume a normal distribution for measurement error. Also, these statements assume a constant CSEM, which is a characteristic of the ACT score scales by design. Note that approximately 36 scale score points were needed so that 68% confidence intervals for scale scores could be created by subtracting and adding two points. The intention was to create a score scale that would discourage users from overinterpreting the meaningfulness of small score differences.

6.1.3 Equipercentile Equating

New forms of the ACT tests are developed each year. Though each form is constructed to adhere to the same content and statistical specifications, the forms may differ slightly in difficulty. To control for these differences, new forms are equated to an older form with an established relationship between number of items answered correctly and 1–36 scale scores. As a result of equating, scale scores reported to examinees have the same meaning across all test forms and test dates.

A carefully selected sample of examinees from a national test date is used as the sample in a random-groups equating design. The examinees in the equating sample are administered a spiraled set of forms including new forms and one anchor form that was equated to previous forms. The forms are spiraled such that randomly equivalent groups of more than 2,000 examinees take each form.

Scores on the new forms are equated to the anchor form score scale using equipercentile equating methodologies. In equipercentile equating, a score on Form X and a score on Form Y are considered equivalent if they are associated with the same percentile rank for the randomly equivalent groups of examinees that took those forms. The equipercentile equating results are smoothed using an analytic method described by Kolen (1984) to establish a smooth relationship between scores on two test forms. The equivalent scores are then rounded to integers. The conversion tables resulting from this process are used to transform raw scores on the new forms to the 1–36 scale scores reported to students.

The above discussion focused on the equating of the four multiple-choice tests of the ACT. Other reported scores that are combinations of multiple test scores are not equated directly. These scores—including the Composite, STEM, and ELA scores—are each a rounded average of the scale scores from two or more tests. More information on these scores is provided in *Chapter 5*. The Composite, STEM, and ELA scores are also comparable across forms because the scores used to compute them have been equated.

6.1.4 Equating for Reporting Category Readiness Ranges and the Understanding Complex Texts Indicator

As described in *Chapter 3*, ACT items are classified into reporting categories that describe specific groups of skills associated with college and career readiness. Student performance on the items in a reporting category is reported on a percentage correct scale, and that score may

fall within an ACT Readiness Range, which indicates the score range expected of students who met or exceeded the corresponding ACT College Readiness Benchmark (see *Chapter 5* for a detailed description of reporting category scores). The ACT Readiness Range can vary across forms due to differences in difficulty and number of items. What follows is the procedure for identifying ACT Readiness Ranges.

To determine the lower bound of a Readiness range, student data are used to create a predictive relationship between ACT scale scores and percentage correct scores in a reporting category. Using that relationship, the lower bound is set as the percentage correct score expected of a student who just met the corresponding ACT College Readiness Benchmark (e.g., 18 on the English test, 22 on the mathematics test, etc.). For example, a Readiness range is developed for each of the three English reporting categories. For the first reporting category—Production of Writing—linear regression is used to estimate a predictive relationship between 1–36 English scale scores and percentage correct scores on the items associated with the Production of Writing reporting category. This relationship is then used to identify the percentage correct score for the reporting category corresponding to the ACT College Readiness Benchmark on the overall English test (18). Students with percentage correct scores at or above the lower bound are considered within the ACT Readiness Range. The upper bound of each ACT Readiness Range corresponds to answering all questions in that reporting category correctly. The same process is repeated to determine Readiness ranges for the other two English reporting categories and the reporting categories of the other multiple-choice tests.

Items on the ACT reading test may be further classified as Understanding Complex Text (UCT) items, which means that they require students to identify the central meaning of complex texts at the level needed to succeed in college courses with higher reading demands. Student performance on UCT items is reported according to three performance levels: Below Proficient, Proficient, or Above Proficient (see *Chapter 5* for a detailed description of the UCT indicator). Proficient indicates that a student has at least a 50% chance of earning a B or higher in seven types of courses (American History, Literature, Other History, Other Natural Science, Physics without Calculus, Sociology, and Zoology) at a typical postsecondary institution, and Above Proficient indicates that a student has at least a 67% chance of earning a B or higher.

As described by Allen, Bolender, Fang, Li, and Thompson (2016), the score ranges corresponding to the three performance levels were initially established as 0–8 for Below Proficient, 9–12 for Proficient, and 13–16 for Above Proficient. However, the number correct scores defining the boundaries between the performance levels can vary across ACT reading forms due to differences in difficulty and number of items. The UCT number correct scores on new reading forms are equated to the original 0–16 scale with the same equipercentile methods used to equate the full multiple-choice tests. After that, the cut scores (9 for Proficient, 13 for Above Proficient) are applied to generate UCT indicators for new reading forms.

6.2 Scaling and Equating of the ACT Writing Test for ACT ELA Score Calculation

ACT began reporting English Language Arts (ELA) scores in September 2015 when the current ACT writing test was launched. A 1–36 score scale was introduced for the current ACT writing test at its launch, and the ELA score is calculated as the rounded average of the English, reading, and writing 1–36 scale scores. Since September 2016, when the 2–12 rounded average domain scores replaced the 1–36 scores for the ACT writing test score reporting, the 1–36 writing scale has solely been used for calculating ELA scores.

In fall 2014, the 1–36 writing scale was constructed based on data from the first special field test study of the current writing test prompts. After evaluating all prompts administered in the special study, one prompt was selected to be the base prompt. This base prompt was used to establish the 1–36 scale for writing. To obtain the base prompt raw-to-scale score conversion, percentile ranks of all raw score points (i.e., the sum of the four domain scores) were calculated. Then the corresponding *z* scores from a standard normal distribution were obtained for these percentile ranks. The *z* scores were then linearly transformed to cover the whole score range of 1–36. Finally, a seventh-degree polynomial regression of the unrounded scale scores on the raw scores was used to slightly smooth the conversion prior to rounding to integer scale scores to obtain the final raw-to-scale score conversion for the base form.

As described in *Chapter 2*, the comparability of the 2–12 writing test scores across forms is maintained by the prompt selection procedures. Prompts are selected to ensure that the 2–12 writing test scores are comparable no matter which prompt the student takes, but that process does not ensure that the prompts are also strictly comparable for the sum of the four domain scores (on an 8–48 scale). Equating is used to adjust for slight differences in prompt difficulty for the sum of the domain scores that may remain after the writing prompt selection process. The same methodology for equating the multiple-choice ACT tests is used for equating each prompt and obtaining the 1–36 writing scale scores: equipercentile equating with post-smoothing under the random groups design. This process ensures year-to-year comparability of the ELA scores. The ELA score is intended to be a more reliable measure of student ability than the ACT writing test score, which is based on a student's response to a single prompt.

6.3 Reliability and Measurement Error

The potential for some degree of inconsistency or error is inherent to the measurement of any cognitive characteristic. An examinee administered one form of a test on one occasion and a second, parallel form on another occasion may earn somewhat different scores on the two administrations. These differences might be due to the examinee or the testing situation, such as differential motivation or differential levels of distractions during the two administrations. These differences may also result from attempting to estimate the examinee's level of skill in a broad domain from a relatively small sample of items. In this chapter, a set of statistics is provided that quantifies the reliability, measurement error, and classification consistency of the ACT test scores.

6.3.1 Reliability and Standard Error of Measurement for ACT Test Forms

Reliability coefficients quantify the level of consistency in test scores across repeated administrations. They range from zero to one, with values near one indicating high consistency and those near zero indicating little or no consistency. Reliability coefficients are usually estimated based on a single test administration by calculating the inter-item covariances. Such coefficients are referred to as estimates of internal consistency reliability. Coefficient alpha (Cronbach, 1951), which is one of the most widely used estimates of internal consistency reliability, was computed for the ACT tests. Coefficient alpha can be computed using the following formula

$$\alpha = \frac{k}{k-1}\left(1 - \frac{\sum_{i=1}^{k} s_i^2}{s_x^2}\right),$$

where *k* is the number of test items, s_i^2 is the sample variance of the *i*th item scores, and s_x^2 is the sample variance of the observed total raw scores.
Coefficient alpha provides reliability estimates for number correct scores. For scale scores, a different reliability estimate (r_t) is obtained using the following formula

$$r_t = 1 - \frac{SEM_t^2}{s_t^2},$$

where SEM_t is the estimated scale score standard error of measurement, and s_t^2 is the sample variance of the observed scale score for test *t*. The standard error of measurement (SEM) summarizes the amount of error or inconsistency in scores on a test. Scale score reliability coefficients and SEMs were estimated using a four-parameter beta compound binomial model as described in Kolen, Hanson, and Brennan (1992). One input to this calculation was an estimate of the relative error variance for a generalizability study with a person × (items: content) design. Note that relative error variance concerns the reliability of test scores for rank ordering examinees. Reported reliability coefficients would have been slightly lower (by 0.01–0.03) using absolute error variance, which concerns the reliability of classifying students as attaining or not attaining a certain score. If measurement error has a normal distribution, true scale scores for about two-thirds of the examinees are within plus or minus one SEM from their reported scale scores.

Reliability and SEM for the ACT Test Scores

Scale score reliability estimates and SEM for the four ACT multiple-choice tests (English, mathematics, reading, and science), Composite, STEM, and ELA scores are provided in Table 6.1. These values were calculated based on operational test data from the test forms administered in the 2024–2025 academic year in the Wisconsin state-sponsored administration of the ACT. By design, the SEM should be about 1 for the Composite score and about 2 for the subject tests. Reliability and SEM values were fairly consistent across forms.

Table 6.1. Scale Score Reliability and SEM for the Wisconsin Spring 2025 ACT Test Scores

Test	# of Items	Standard Form		Accommodated Form	
		Reliability	SEM	Reliability	SEM
English	75	0.93	1.70	0.79	1.76
Mathematics	60	0.91	1.77	0.74	1.70
Reading	40	0.89	2.15	0.78	2.26
Science	40	0.85	2.16	0.67	2.26
Composite	215	0.97	0.98	0.91	1.01
ELA	116	0.94	1.41	0.88	1.44
STEM	100	0.93	1.40	0.81	1.41

Note: Information in this table can also be found in *Appendix Table 13*.

Reliability and SEM for ACT Reporting Category Scores

Raw score reliability (coefficient alpha) and SEM were also calculated for the ACT reporting categories. These values, provided in Table 6.2, were calculated using operational test data from forms administered in the 2024–2025 academic year in the Wisconsin state-sponsored administration of the ACT. For some of the reporting categories, particularly those with very few items, the reliability was low. However, reporting category scores are not intended for use in making high-stakes decisions about students. Rather, they are intended to guide instruction and help identify students' strengths and weaknesses.

Table 6.2. Raw Score Reliability and SEM for the Wisconsin Spring 2025 ACT Reporting Categories

Test/reporting categories	Standard			Accommodated		
	# of items	Reliability	SEM	# of items	Reliability	SEM
English						
Production of Writing	22	0.82	2.01	22	0.61	2.09
Knowledge of Language	12	0.78	1.46	12	0.43	1.53
Conventions of Standard English	41	0.88	2.78	41	0.62	2.78
Mathematics						
Preparing for Higher Math	36	0.85	2.57	36	0.56	2.57
Number & Quantity	6	0.52	1.09	6	0.25	1.02
Algebra	8	0.66	1.18	8	0.20	1.17
Functions	8	0.47	1.23	8	0.16	1.26
Geometry	8	0.43	1.17	8	0.12	1.28
Statistics & Probability	6	0.56	1.04	6	0.36	0.98
Integrating Essential Skills	24	0.83	2.11	24	0.65	1.99
Modeling	25	0.78	2.17	23	0.60	2.05
Reading						
Key Ideas & Details	21	0.81	1.99	22	0.69	2.06
Craft & Structure	12	0.73	1.49	11	0.47	1.48
Integration of Knowledge & Ideas	7	0.52	1.17	7	0.34	1.13
Science						
Interpretation of Data	16	0.71	1.78	20	0.64	1.96
Scientific Investigation	10	0.59	1.42	10	0.27	1.44
Evaluation of Models, Inferences & Experimental Results	14	0.73	1.62	10	0.49	1.33

Note: Information in this table can also be found in *Appendix Table 15*.

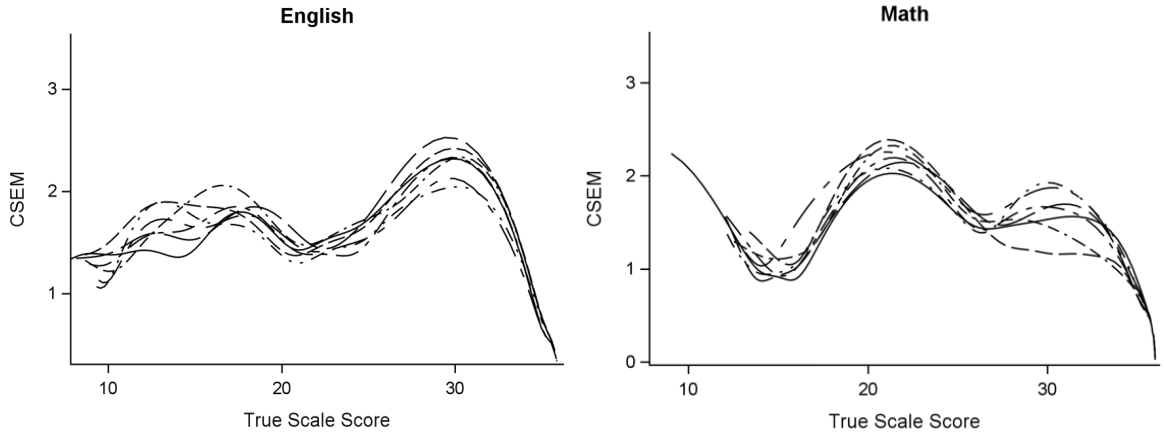
Conditional Standard Errors of Measurement for the ACT Multiple-Choice Test Scores

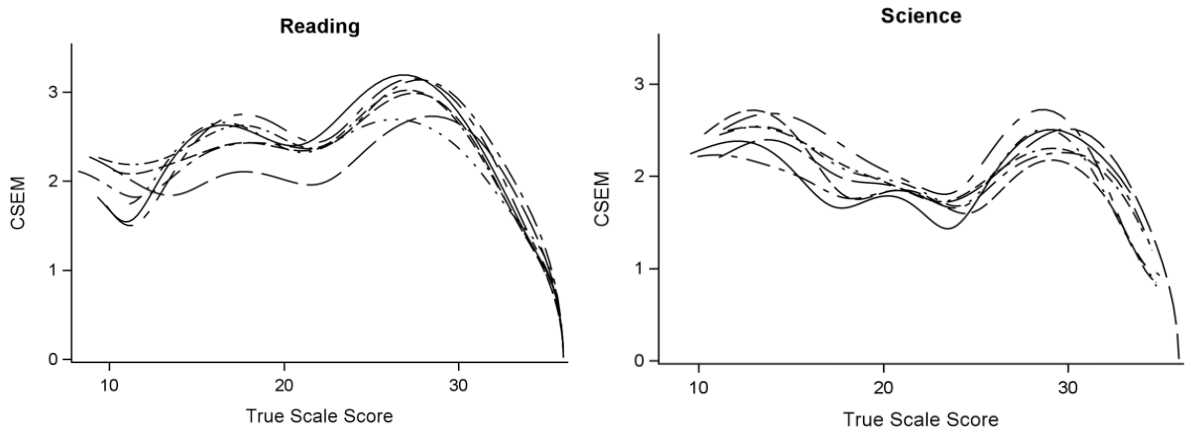
Whereas the SEM indicates average score uncertainty (or imprecision) across the entire score scale, the conditional standard error of measurement (CSEM) quantifies the uncertainty at a particular score. The score scales for the ACT were developed to have approximately constant CSEMs for all true scale scores. This statement implies, for example, that the CSEM for an ACT scale score is approximately the same for low-scoring examinees and high-scoring examinees.

For the ACT, the CSEMs were computed using methods described by Kolen, Hanson, and Brennan (1992). Figure 6.1 presents the CSEMs for the four multiple-choice tests for 7 of the forms administered from September 2024 to July 2025. The CSEM is not plotted for very low scale scores that can be obtained by guessing or random responding. The minimum scale scores at which the CSEM was plotted were chosen such that only an extremely small proportion of examinees would be expected to have a true scale score lower than the minimum plotted score.

The ACT tests were scaled to have an approximately equal CSEM as close to 2 as possible along the score scales. That property is best observed in the science test in Figure 6.1. The CSEMs of the English, mathematics, and reading tests had greater variation along the score scale, but in most of the true scale score range, the CSEM is about 2 or lower. For all tests, the CSEM approaches zero as the true scale score approaches the maximum of 36. For this reason, the CSEM cannot be perfectly constant for all true scale scores.

Figure 6.1. CSEM for Multiple-Choice Test Scores



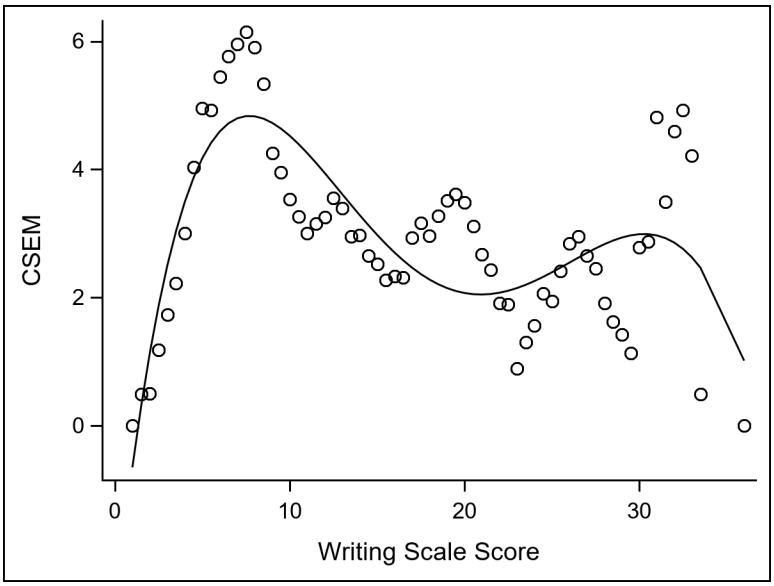


Reliability, CSEM, and Agreement Indices for the ACT Writing Test Scores

Reliability and CSEM for the ACT writing test were estimated using results from a generalizability study. To investigate the properties of the overall writing score and the domain scores, a generalizability study was conducted in fall 2014. The study was separated into three parts, each involving a different pair of schools. Within each pair of schools, two writing prompts were administered. The responses to both prompts were rated by three raters on the four writing domains. The same raters rated both prompts for both schools. Different pairs of prompts and different groups of three raters were used for each pair of schools. This essentially served as three replications of the same study. The estimated variance components for the rater by prompt interaction and the rater by person (or student) interaction were small across all three school pairs. This indicated that raters behaved similarly across prompts and that students received similar evaluations from different raters. In contrast, the estimated variance component for the person by prompt interaction was relatively large for all three pairs of schools. This finding was consistent with results typically observed in the research literature on extended-response assessments. For the average of the domain scores, the generalizability coefficients (reliability-like estimates of score consistency) ranged from 0.61 to 0.77, which are fairly high for a writing assessment. SEMs ranged from 0.84 to 1.10.

Data from the 2019 writing field test study were used to estimate the reliability and SEM for writing scores on the 1–36 scale used for calculating ELA scores. Each student took two different prompts. The data were analyzed using a person by occasion generalizability study design. The individual conditional error variances were fit with a fifth-degree polynomial. The square root of these fitted values is represented by the solid line in Figure 6.2. The average CSEM values, represented by the circles, were calculated by taking the square root of the average conditional error variances at each scale score point. The generalizability coefficient was 0.74 and the scale score SEM was 3.23. This SEM value was used to calculate the ELA reliability and SEM.

Figure 6.2. Average and Fitted CSEMs for ACT Writing Test Scale Scores



Indices of operational rater agreement were calculated based on the forms administered in the 2024–2025 academic year in the Wisconsin state-sponsored administration of the ACT. This included the perfect agreement rate, the perfect plus adjacent agreement rate, and the quadratic weighted kappa coefficient (Table 6.3).

The quadratic weighted kappa coefficient (Cohen, 1968) is a measure of agreement between raters for categorical scores (e.g., 1, 2, 3). It uses weights to account for the relative differences between categories. In the calculation, for example, a 2-point disagreement is weighted more than a 1-point disagreement. The kappa coefficient is a positive number if the observed agreement is larger than the chance level of agreement, with larger numbers representing stronger agreement between two raters. Fleiss, Levin, and Paik (2013) indicated that for most purposes, kappa values larger than 0.75 represent excellent agreement beyond chance, values below 0.40 represent poor agreement beyond chance, and values in between represent fair to good agreement beyond chance. The quadratic weighted kappa coefficients for the ACT writing domain scores ranged from 0.82 to 0.86, indicating good rater agreement.

Table 6.3. Agreement Rates for the ACT Writing Domain Scores for the Wisconsin Spring 2025 Administration

Domain	Agreement Index	Value
Ideas & Analysis	Perfect Agreement	0.73
	Perfect + Adjacent Agreement	>0.99
	Quadratic Weighted Kappa	0.86
Development & Support	Perfect Agreement	0.72
	Perfect + Adjacent Agreement	>0.99
	Quadratic Weighted Kappa	0.85
Organization	Perfect Agreement	0.73
	Perfect + Adjacent Agreement	>0.99
	Quadratic Weighted Kappa	0.85
Language Use & Conventions	Perfect Agreement	0.70
	Perfect + Adjacent Agreement	0.99
	Quadratic Weighted Kappa	0.82

Note: Information in this table can also be found in *Appendix Table 16.*

CSEM for Composite Scores

Assuming that measurement errors on the four ACT multiple-choice tests (English, mathematics, reading, and science) are independent, the CSEM for the unrounded Composite score is

$$s_c(\tau_e, \tau_m, \tau_r, \tau_s) = \frac{\sqrt{\sum_i s_i^2(\tau_i)}}{4}$$

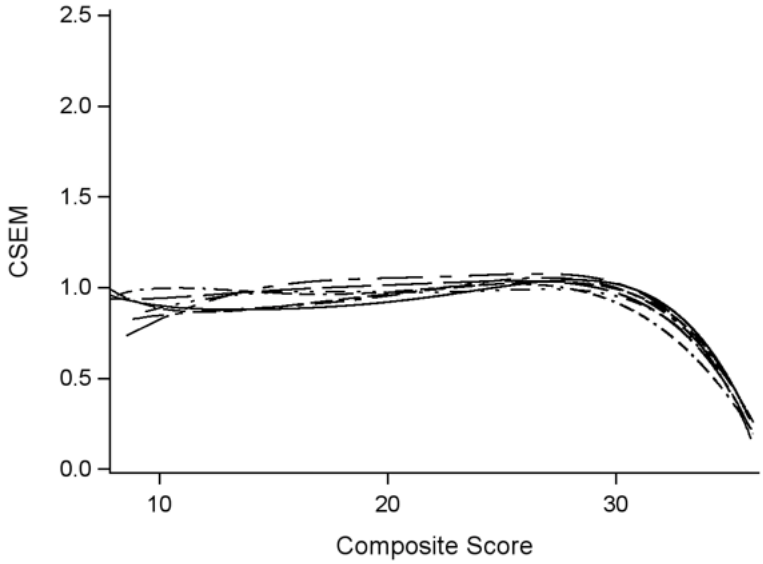
where $s_i(\tau_i)$ is the CSEM for test *i* at true scale score τ_i and *i* = *e*, *m*, *r*, and *s* for English, mathematics, reading, and science, respectively. The $s_i(\tau_i)$ functions are plotted in Figure 6.1. A particular true Composite score can be obtained in a variety of ways (i.e., different combinations of true scale scores on the individual tests could produce the same true Composite score). Consequently, each true Composite score value may correspond to several different values of the CSEM, depending on the combination of true scores on the four tests that produced the true Composite score value.

To produce CSEM plots for Composite scores, the observed proportion-correct scores (i.e., the number of items answered correctly divided by the total number of items) for examinees on the four tests were treated as true proportion-correct scores at which the CSEMs were calculated. For each test, the CSEM was computed for each examinee using the observed proportion-correct score as the true proportion-correct score in the formula for the CSEM (Equation 8 in Kolen, Hanson, & Brennan, 1992). In addition, for each test the true scale score corresponding to the observed proportion-correct score (treated as a true proportion-correct score) was computed (Equation 7 in Kolen, Hanson, & Brennan, 1992). The resulting CSEMs for the four tests were substituted in the equation given above to compute the CSEM for the Composite score. A fifth-degree polynomial regression was used to get a unique CSEM value for each

Composite score for each test form. The CSEMs for the Composite score of 7 test forms administered in September 2024 to July 2025 are plotted in Figure 6.3. The CSEMs of the Composite score were reasonably constant across the score scale.

A limitation of the approach used in producing the CSEM estimates of the Composite score in Figure 6.3 is that they correspond to the unrounded average of the four test scores rather than the rounded average of the four test scores, which is the Composite score reported to examinees.

Figure 6.3. CSEM for Composite Scores



CSEM for STEM and ELA Scores

The CSEMs for the STEM and ELA scores were calculated using the same approach used to calculate the CSEM for the Composite score. Assuming that measurement errors on the four multiple-choice tests are independent, the CSEM for the unrounded STEM score is

$$S_{STEM}(\tau_m, \tau_s) = \frac{\sqrt{\sum_i s_i^2(\tau_i)}}{2}$$

where *i* = *m* and *s* for mathematics and science, respectively. Similarly, the CSEM for the unrounded ELA scores is

$$S_{ELA}(\tau_e, \tau_r, \tau_w) = \frac{\sqrt{\sum_i s_i^2(\tau_i)}}{3}$$

where $s_i(\tau_i)$ is the CSEM for test *i* at true scale score τ_i and *i* = *e*, *r*, and *w* for English, reading, and writing, respectively. The same set of data used to produce the CSEM values for the

Composite score was used to obtain the CSEM values for the STEM scores plotted in Figure 6.4 and the CSEM values for the ELA scores in Figure 6.5.

Figure 6.4. CSEM for STEM Scores

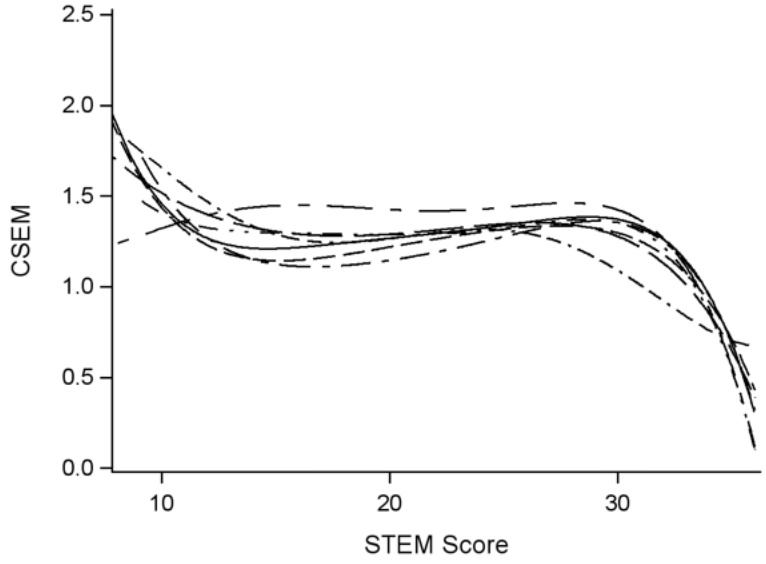
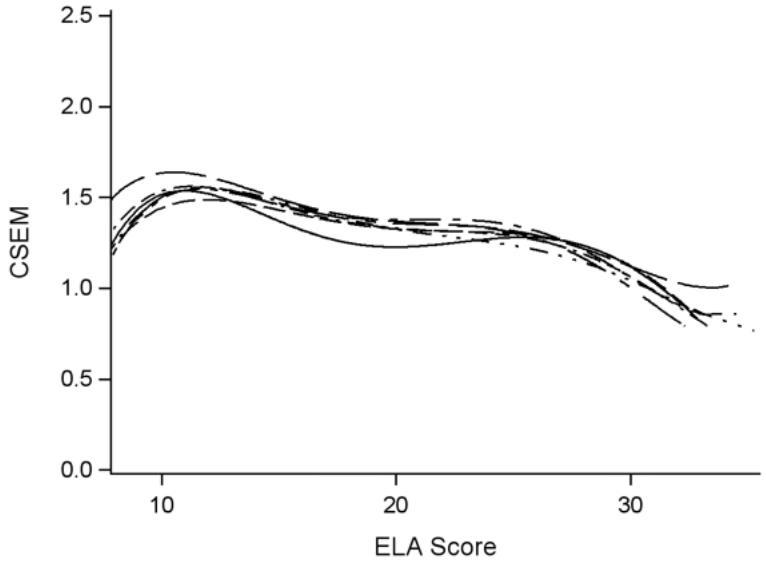


Figure 6.5. CSEM for ELA Scores



6.3.2 Classification Consistency

Classification consistency refers to the extent to which examinees are classified into the same category over replications of a measurement procedure. Because tests are rarely administered twice to the same examinee, classification consistency is typically estimated from a single test administration with strong assumptions about distributions of measurement errors and true scores (e.g., Hanson & Brennan, 1990; Livingston & Lewis, 1995).

Using the method described by Livingston and Lewis (1995), the true score distribution was estimated by fitting a four-parameter beta distribution. The expected conditional distribution of scores, given the true score, is a binomial distribution. With the assumption of independent errors of measurement, the probabilities that a student would be classified into each pair of categories were computed, given the true score. The conditional results were then aggregated over the true score distribution to get a contingency table containing probabilities of a student receiving scores from two administrations that fall into any combination of categories. The estimated classification consistency index for the whole group is the sum of the values on the diagonal of the contingency table, which represent the probabilities of being classified in the same category on two separate administrations.

Classification Consistency for the ACT Multiple-Choice Test, STEM, and ELA Scores

Classification consistency values were computed using data from the forms administered in the 2024–2025 academic year in the Wisconsin state-sponsored administration of the ACT. Classification was based on the Wisconsin ACT Performance level cut scores that were developed in a standard setting workshop conducted in the summer of 2024. Additional information regarding the workshop can be found on the WPDI website in the *2024 Wisconsin ACT Performance Level Standard Setting Report*. Table 6.4 summarizes the agreements between the operational test classifications, showing the percentages of students who would be placed in the same achievement level on two equivalent administrations of the test. Classification consistencies were also computed for subgroups, provided in Table 6.5.

- ELA
  - Approaching: 14
  - Meeting: 19
  - Advanced: 26
- Mathematics
  - Approaching: 16
  - Meeting: 19
  - Advanced: 27
- Science
  - Approaching: 17
  - Meeting: 21
  - Advanced: 26

Table 6.4. Classification Consistency for the Wisconsin Spring 2025 Performance Level Cut Scores

Test	Number of Items	Classification Consistency	
		Two-level	Four-level
ELA	116	0.90	0.79
Mathematics	60	0.88	0.76
Science	40	0.85	0.69

Note: Information in this table can also be found in *Appendix Table 18*.

Table 6.5. Classification Consistency for the Wisconsin Spring 2025 ACT Readiness Ranges

	Standard form (n=13,829)				Female (n=6,476)			
	Two Levels		Four Levels		Two Levels		Four Levels	
	Agreement	Kappa	Agreement	Kappa	Agreement	Kappa	Agreement	Kappa
ELA	0.895	0.790	0.755	0.656	0.891	0.777	0.751	0.642
Mathematics	0.880	0.757	0.660	0.532	0.872	0.740	0.653	0.516
Science	0.849	0.694	0.587	0.439	0.837	0.667	0.582	0.425
	Male (n=6,766)				**African-American (n=707)**			
	Two Levels		Four Levels		Two Levels		Four Levels	
	Agreement	Kappa	Agreement	Kappa	Agreement	Kappa	Agreement	Kappa
ELA	0.899	0.797	0.758	0.665	0.921	0.746	0.768	0.639
Mathematics	0.884	0.769	0.669	0.548	0.924	0.659	0.739	0.476
Science	0.857	0.712	0.597	0.456	0.890	0.524	0.671	0.396
	Asian (n=596)				**American-Indian (n=104)**			
	Two Levels		Four Levels		Two Levels		Four Levels	
	Agreement	Kappa	Agreement	Kappa	Agreement	Kappa	Agreement	Kappa
ELA	0.909	0.818	0.749	0.656	0.903	0.759	0.785	0.673
Mathematics	0.887	0.773	0.650	0.525	0.884	0.664	0.718	0.517
Science	0.862	0.718	0.560	0.405	0.819	0.487	0.629	0.413
	Hispanic (n=1,990)				**White (n=9,030)**			
	Two Levels		Four Levels		Two Levels		Four Levels	
	Agreement	Kappa	Agreement	Kappa	Agreement	Kappa	Agreement	Kappa
ELA	0.897	0.773	0.751	0.643	0.891	0.774	0.754	0.646
Mathematics	0.893	0.724	0.669	0.491	0.866	0.730	0.651	0.518
Science	0.863	0.635	0.618	0.424	0.836	0.671	0.570	0.416
	Two or more races (n=710)							
	Two Levels		Four Levels					
	Agreement	Kappa	Agreement	Kappa				
ELA	0.900	0.801	0.758	0.666				
Mathematics	0.892	0.773	0.654	0.518				
Science	0.862	0.713	0.606	0.457				
	English learner (n=462)				**Accommodated form (n=1,772)**			
	Two Levels		Four Levels		Two Levels		Four Levels	
	Agreement	Kappa	Agreement	Kappa	Agreement	Kappa	Agreement	Kappa
ELA	0.900	0.756	0.758	0.646	0.968	0.748	0.863	0.664
Mathematics	0.901	0.741	0.670	0.495	0.956	0.722	0.757	0.413
Science	0.874	0.644	0.627	0.426	0.944	0.697	0.728	0.399

Note: Information in this table can also be found in *Appendix Table 19*.

Classification Consistency for ACT Understanding Complex Texts Indicator

Classification consistency was also computed for two other indicators provided on ACT score reports. The first indicator is Understanding Complex Texts (UCT). Across seven of the forms

administered from September 2024 to July 2025, the classification consistency ranged from 0.62 to 0.66, which was moderately high considering the number of items that contribute to UCT scores and the number of performance levels. Specifically, the number of UCT items ranged from 15 to 21 across these seven forms, and the percentages of students classified as Below Proficient, Proficient, and Above Proficient were 39%, 26%, and 35%, respectively.

Classification Consistency for Progress Toward ACT NCRC Indicator

The second indicator, Progress Toward the ACT National Career Readiness Certificate® (ACT NCRC®), had classification consistency values ranging from 0.79 to 0.81 across seven of the forms administered from September 2024 to July 2025. These values are quite high considering that there are four performance levels for the ACT NCRC, as shown in Table 6.6. Note that the classification consistency index is an indication of the stability of the Progress Toward ACT NCRC indicator if different ACT test forms were taken and is not an indication of the accuracy of the classification compared with students' actual NCRC attainment. See *Chapter 5* for more information about the Progress Toward the ACT National Career Readiness Certificate indicator.

Table 6.6. Composite Score Ranges for the ACT NCRC Levels

ACT NCRC Level	Composite Score Range
Unlikely to earn an ACT NCRC	1–12
Most Likely to obtain a Bronze level on the ACT NCRC	13–16
Most Likely to obtain a Silver level on the ACT NCRC	17–21
Most Likely to obtain a Gold level on the ACT NCRC	22–26
Most Likely to obtain a Platinum level on the ACT NCRC	27–36

6.4 Mode Comparability for Online Testing

6.4.1 Overview of ACT Online Test Administration

ACT launched a pilot study for the first-ever online administration of a national undergraduate college admission exam in April 2014. In this study, the ACT was administered to approximately 4,000 students at 80 test sites, and college reportable scores were provided.

In April 2015, online testing was expanded to a limited number of test sites in the United States, with more than 6,000 students receiving college reportable scores. Online testing for the ACT was then offered to all state and district test sites starting in 2016, and it will continue to be offered going forward. Beginning in September 2018, all international testing occurs online.

As of spring 2020, the ACT may be administered on paper or online for state and district testing and online only for international students. At present, a very small number of students eligible for the screen reader accommodation take the ACT online during national administrations. State and district online testing is delivered during multiple testing windows, each of which provides test access over a short period. Online administration of the ACT follows the administration guidelines established for paper testing wherever appropriate.

6.4.2 Online Platform and Capabilities

ACT collaborated with Pearson to design the TestNav platform architecture for the ACT online test delivery system. Test centers can use this test delivery system across multiple device types, including laptop and desktop computers running operating system such as macOS, Microsoft Windows, and Chrome OS. ACT continually updates the minimum test delivery system requirements to ensure compatibility with test delivery technology.

The most current technical requirements for taking the ACT online are available at *http://www.act.org/content/dam/act/unsecured/documents/TechnicalGuidefortheACTTakenOnline.pdf*.

Similarly, ACT worked with PSI to customize the ATLAS Cloud testing platform for international ACT testing online. International test centers can administer the ACT on desktop and laptop computers running Microsoft Windows or macOS. The current technical requirements for taking the ACT online via ATLAS Cloud are available at *https://global.act.org/content/global/en/products-and-services/the-act-non-us/international-cbt/technical-requirements.html*.

6.4.3 Comparability of Scores between Online and Paper Testing

ACT maintains the comparability of scores between online and paper administrations of the ACT test by conducting mode comparability studies and subsequent online form equating. Initial online forms were linked to paper forms through equating methodologies based on data gathered in special mode comparability studies where both paper and online forms were administered. For state and district testing, subsequent online forms are equated to the online base forms through online test equating studies. ACT uses the same data collection designs and test equating procedures to link online scores to paper scores and to equate the online forms as it uses to equate the ACT paper test forms. For international testing, IRT true-score equating is employed to generate raw score to scale score conversion tables appropriate for online testing. These procedures are described in detail in *Section 6.1.5* in the ACT Technical Manual.

6.4.4 ACT Online Timing and Mode Comparability Studies

As part of the initial development process of delivering the ACT online, ACT conducted several special studies to evaluate the comparability of scores between online and paper administrations before the official launch of the ACT online tests, including a timing study in fall 2013, a mode comparability study in spring 2014, and a second mode comparability study in spring 2015. In 2018, another mode comparability study was conducted in preparation for online testing for the ACT international program. Then, between 2019 and 2020, a series of three mode comparability studies were conducted to support current and future use of the TAO platform for online ACT delivery.

All studies used a randomly equivalent groups design. That is, students were randomly assigned to take the test under different timing conditions in the online timing study and were randomly assigned to take the paper or online test in the mode studies. ACT reevaluated timing

recommendations from the timing study in the subsequent mode study, which resulted in a modification of the initial timing decisions for the online administration. The updated timing for online administration was then implemented in the 2015 mode study. Provided below are brief summaries of these studies. See Li, Yi, and Harris (2017) and Steedle, Pashley, and Cho (2020) for more details.

Fall 2013 Timing Study

The purpose of the timing study was to evaluate whether the online administration of the ACT would require different time limits from the paper administration. The four multiple-choice tests were administered online to approximately 3,000 examinees, with each examinee taking one test. Students were randomly assigned to take the test under one of three timing conditions: the current standard paper time limit (i.e., 45, 60, 35, and 35 minutes for English, mathematics, reading, and science tests, respectively), the current time limit plus 5 minutes, and the current time limit plus 10 minutes. At the end of the test, the students were also given a survey with questions regarding their testing experience, including whether they felt they had enough time to finish the test. Students in this study did not receive college reportable scores.

Item and test level scores, item omission rates, item and test latency information, and student survey results were analyzed using a variety of methods, both descriptive and inferential. Because the timing study had only online test administrations, a matched sample based on total score distributions was also created from operational paper testing data of the same test form. Item mean scores (i.e., item p-values) and omission rates were compared between the timing study sample and the matched sample.

Results from various analyses suggested that the online reading and science tests under the current standard timing condition might be more speeded than paper testing. For example, compared with the matched operational paper sample, the average number of items omitted was higher for the timing study sample for all subject tests under the current standard paper testing timing condition. The timing study sample also had lower item p-values for the last few items than the matched sample, especially for reading and science. In addition, among the students who responded to the survey questions, about half either disagreed or strongly disagreed with the statement that they had enough time to complete the reading and the science tests.

However, findings from the timing study might have been confounded with issues of low motivation and unfamiliarity with the online testing format. For example, even though an online tutorial was provided to students before they took the tests, the posttest survey indicated that less than half of the students made use of this resource, with an even lower percentage for students who took the reading and the science tests. After the results of various analyses were evaluated from different perspectives, ACT decided to tentatively increase online testing time for the reading and science tests by 5 minutes. Also, ACT planned a subsequent mode comparability study to continue evaluating the timing issue.

Spring 2014 Mode Comparability Study

To gather additional information about the differences between online and paper testing modes and to learn about administration issues, ACT conducted a mode comparability study in an operational testing environment wherein participating students received college reportable scores. The purposes of the mode comparability study were to:

(1) investigate the comparability of the scores from the two testing modes;

(2) obtain interchangeable scores across modes for operational score reporting;

(3) reevaluate the timing decisions for the online administration of the reading and science tests; and

(4) gain insights into the online administration process.

Students participating in the spring 2014 study were randomly assigned to take one of the three forms administered in the study (one paper and two online). After the administration, survey questions were sent to students who participated in the study to gather their comments and feedback on their testing experiences.

More than 7,000 students from about 80 high schools across the country signed up for this study. Data were cleaned based on reviews of the proctor comments, phone logs, irregularity reports, latency information, and an evaluation of the random assignment. Students with invalid scores and test centers with large discrepancies in form counts across modes were excluded from further analyses.

Using data from paper and online forms comprising the same items, analyses were conducted to investigate mode comparability from two perspectives: construct equivalency and score equivalency. Construct equivalency was examined by comparing the dimensionality and factor loadings and by examining differential item functioning (DIF) between online and paper items. Score equivalency was examined in terms of the similarity of test score distributions between the two modes, such as means, standard deviations, and relative cumulative frequency distributions. For the English, mathematics, reading, and science tests, the similarity of item score distributions, such as the item p-values, item response distributions across the different options for each item, and item omission rates were compared. In addition, measurement precision (i.e., reliability and conditional standard errors of measurement) was compared across modes, and the item latency information for the online test items was also examined.

Results revealed little difference between the two modes in terms of test reliability, correlations among tests, effective weights, and factor structures. However, item scores and test scores tended to be higher and omission rates tended to be lower for the online group compared to the paper group, especially for the reading test but also for the science and English tests. Equating methodology was applied to each of the four multiple-choice tests to adjust for mode differences, which ensured that the college reportable scores of students participating in the mode comparability study were comparable to national examinees, regardless of the testing

mode. Based on the findings from the spring 2014 mode comparability study, ACT decided to eliminate the extra 5 minutes for the online reading and science tests. Another mode comparability study was conducted in spring 2015 with the revised timing decisions for online testing.

Spring 2015 Mode Comparability Study

The mode comparability study in spring 2015 was to further examine the comparability between online and paper scores and the impact of eliminating the extra 5 minutes for the reading and science online tests. More than 4,000 students from more than 40 schools signed up to participate in this study. One paper form and two online forms were administered. In addition, students who participated in the 2015 study all took the redesigned ACT writing test, which was to be launched in fall 2015. The spring 2015 study followed the same design as the 2014 study, and similar analyses were conducted for the four multiple-choice tests.

Results showed that students performed similarly across modes on the science test but still higher on the online reading test even without the extra 5 minutes. To a similar degree, online English scores were higher than paper English scores. To adjust for mode effects, equating methodology was applied to produce comparable scores regardless of the testing mode. For the two prompts included in the writing mode study, students performed similarly across modes on one prompt but differentially on the other, with online scores higher than paper scores on average.

Summary of TestNav Studies

The ACT online timing study and the two mode comparability studies all used the gold standard of research design: random assignment to timing or mode conditions. The two mode comparability studies, one with initial timing decisions and one with the final timing decisions for the online administration, were both conducted in an operational testing environment where student motivation was expected to be high.

Whereas the analyses indicated comparability between modes in terms of the construct equivalence and measurement precision, slight differences were observed on item-level and test-level statistics. Under the final online timing conditions, the largest mean differences between modes were observed for the reading and English tests, which were approximately one scale score point (or an effect size of 0.18 or 0.17 standard deviations, respectively). Considering that the standard error of measurement of the test is about two scale score points, the apparent mode effect was small. However, due to the high-stakes uses of the test scores, a systematic score difference of even one score point may have practical impact.

Therefore, ACT used test equating methodology to ensure comparability of scores between paper and online administrations. To maintain ACT score comparability regardless of testing mode, online test forms administered for state and district testing are equated to the base online form, which was linked to paper forms through the spring 2015 mode study.

Chapter 7
Validity Evidence for the ACT Tests

According to the Standards for Educational and Psychological Testing (AERA, APA, & NCME, 2014), “Validity refers to the degree to which evidence and theory support the interpretations of test scores for proposed uses of tests” (p.11). Arguments for the validity of an intended inference made from a test score may contain logical, empirical, and theoretical components. A distinct validity argument is needed for each intended use of a test score.

The potential interpretations and uses of ACT® test scores are numerous and diverse, and each needs to be justified by a validity argument. This chapter describes content, construct, or criterion validity evidence for five of the most common interpretations and uses: measuring students’ educational achievement in particular subject areas, making college admission decisions, making college course placement decisions, evaluating students’ likelihood of success in the first year of college and beyond, and using ACT scores to assist with program evaluation.

7.1 Using ACT Scores to Measure Educational Achievement

The ACT tests are designed to measure students’ problem-solving skills and knowledge in particular subject areas. The usefulness of ACT scores for this purpose provides the foundation for validity arguments for more specific uses (e.g., course placement). This section comprises eleven subsections and provides validity evidence for using ACT test scores to measure students’ educational achievement. The first subsection summarizes content validity evidence supporting the interpretation of ACT scores as a measure of educational achievement. The second covers evidence from cognitive lab studies. The next five subsections focus on relating high school coursework, grades, end-of-course exam scores, and noncognitive factors to ACT scores and ACT Benchmark attainment. The eighth subsection focuses on understanding subgroup differences on the ACT. The ninth subsection focuses on the relationships between test preparation activities and ACT performance. The tenth subsection addresses the use of ACT scores for measuring educational achievement for gifted and talented programs. The final subsection describes validity evidence related to the interpretation of scores for examinees who use available English learner supports during the test.

7.1.1 Content-Oriented Evidence for ACT Scores

The guiding principle underlying the development of the ACT is that the best way to predict success in college is to measure as directly as possible the degree to which each student has developed the academic skills and knowledge that are important for success in college. Tasks presented in the tests must therefore be representative of scholastic tasks. They must be intricate in structure, comprehensive in scope, and significant in their own right, rather than narrow or artificial tasks that can be defended for inclusion in the tests solely on the basis of their statistical correlation with a criterion. Thus, content-related validity is particularly significant

in this context. In other words, assessment tasks must be designed to match the content and cognitive demands of the associated academic domain.

The ACT tests contain a proportionately large number of complex problem-solving exercises and few measures of narrow skills. The tests are oriented toward major areas of college and high school instructional programs. Thus, ACT scores and skill statements based on the ACT College and Career Readiness Standards are directly related to student educational progress and can be readily understood and interpreted by instructional staff, parents, and students.

As described in Chapters *2* and *3*, the test development procedures include an extensive review process, with each item being critically examined at least 16 times. Detailed test specifications have been developed to ensure that the test content is representative of current high school and college curricula. All test forms are reviewed to ensure that they match these specifications. Hence, there is an ongoing evaluation of the content validity of the tests during the development process.

The standardization of the ACT tests is also important to their proper use as measures of educational achievement. Because ACT scores have the same meaning for all students, test forms, and test dates, they can be interpreted without reference to these characteristics.[3] The courses students take in high school and the grades they earn are also measures of educational achievement, but these variables are not standardized because course content varies considerably among schools and grading policies vary among instructors. Therefore, while high school courses taken and grades earned are measures of educational achievement, their interpretation should properly take into account differences in high school curricula and grading policies. ACT scores, because they are standardized measures, are more easily interpreted for the purpose of comparing students than are courses taken and grades earned.

7.1.2 Evidence from Cognitive Lab Studies

Cognitive lab studies involve think-aloud protocols, wherein examinees speak their thoughts while responding to assessment items. This is often followed by structured interviews to further probe examinees' cognitive processes. The goals of cognitive lab studies are typically twofold: to improve item accessibility by identifying construct-irrelevant barriers to responding correctly (e.g., points of confusion) and to evaluate whether items elicit cognitive processes consistent with the construct and depth of knowledge intended to be measured by the items. When items elicit the intended cognitive processes, this confirms alignment of the items to content standards and supports the validity of score interpretations for intended uses such as measuring educational achievement.

Since 2017, ACT has conducted several cognitive lab studies and follow-up analyses. Evidence collected through think-aloud protocols for ACT English and reading items largely supported two

[3] ACT scores obtained before October 1989, however, are not directly comparable to scores obtained in October 1989 or later. A new version of the ACT was released in October 1989 (the "enhanced" ACT). Although scores on the current and former versions are not directly comparable, approximate comparisons can be made using a concordance table developed for this purpose (American College Testing Program, 1989).

overarching claims: the test items required targeted skills found in the ELA standards to obtain the correct answer, and the items did not involve construct-irrelevant factors. Most English items included in the study required students to use the context of the sentence and whole passage to answer correctly. Additionally, students showed evidence that they applied knowledge about grammar and mechanics conventions and discourse knowledge such as whether certain transition words “make more sense” or “flows better” (quotes from study participants). For most reading items, students returned to the passage and applied strategies like skimming, underlining, and summarizing to locate and process relevant information that they used to answer questions. The items clearly required students to use passage evidence— a core component of ELA reading standards—which was illustrated by the way that, for most items, students searched the passage for evidence in order to eliminate options and cited details in the passage as rationales for their answer choice.

During the 2019–2020 school year (prior to the pandemic shutdown of schools), ACT conducted cognitive lab studies that including eye tracking, a think-aloud protocol for reading, surveys, and guided interviews for reading, science, and math. For the reading section, students generally did not have difficulty completing the two passages in the allotted time. Participants identified as high scoring (based on a separate test administration) tended to use more efficient gaze paths (i.e., eye movement patterns) and were able to clearly articulate why they selected specific answers with references to the passage.

For the math cognitive lab studies, eye tracking data provided evidence of cognitive processes. For simple procedural questions, such as finding the median of a data set, all participants who answered correctly showed gaze paths consistent with the skill required (e.g., reordering the data). As expected, high-scoring students exhibited vision paths consistent with one of the optimal solution paths based on the skill map of the question. Additionally, high scorers did not scan the page repeatedly or require multiple rereads of the stem for more difficult items, which was not the case for low and middle scorers. Problems that required complex problem solving showed significant differences between the high scorers and low and middle scorers, which was consistent with high school and postsecondary instructor evaluations of problem solving in the ACT National Curriculum Survey. Timing for items was consistent with skill identification, with easy items taking less time than medium-difficulty items, which took less time than difficult items.

During the science cognitive lab studies, low and middle scorers were more likely to spend time looking at the wrong graphic, particularly when the information needed was not in the first graphic presented. This was true even when the question stem specified which graphic was relevant to the question. Similar to the reading passages with graphics, students required more time and had more return visits in their gaze path for less familiar graphics (e.g., multiple line graphs, phase diagrams, and process diagrams) than for bar graphs and tables. Students answering items correctly generally followed gaze paths indicating the application of skills as described in the content target of the item. Many students spent significant time rereading the stem or response options multiple times, which could have indicated difficulty decoding the task. Students cited familiarity with the overall topic as making a passage easier. In general, cognitive lab study results have been consistent with the claim that ACT items elicit evidence of the skills

they are intended measure. ACT plans to continue such studies, particularly when considering use of new item types or item assessment delivery platforms.

References

ACT. (1989). *Using concordance tables.* Iowa City, IA: ACT.

ACT. (2007). *ACT National Curriculum Survey® 2007.* Iowa City, IA: ACT.

ACT. (2009a). *ACT National Curriculum Survey® 2009.* Iowa City, IA: ACT.

ACT. (2009b). *The ACT Interest Inventory technical manual*. Iowa City, IA: ACT.

ACT. (2010). *Mind the gaps: How college readiness narrows achievement gaps in college success*. Iowa City, IA: ACT.

ACT. (2013a). *ACT National Curriculum Survey® 2013.* Iowa City, IA: ACT.

ACT. (2013b). *ACT profile report—national: Graduating class 2013—national.* Iowa City, IA: ACT.

ACT. (2015). *Linking the current and former ACT writing tests*. Iowa City, IA: ACT.

ACT. (2016a). *ACT National Curriculum Survey® 2016.* Iowa City, IA: ACT

ACT. (2016b). *ACT national profile report*. Iowa City, IA: ACT.

ACT. (2020). *ACT National Curriculum Survey® 2020.* Iowa City, IA: ACT.

ACT. (2025). *ACT Technical Manual*. *https://www.act.org/content/dam/act/unsecured/documents/ACT_Technical_Manual.pdf*

Allen, J. (2013). *Updating the ACT College Readiness Benchmarks*. Iowa City, IA: ACT.

Allen, J. (2015). *How many students use ACT State and District testing to take their sole ACT test?* Iowa City, IA: ACT.

Allen, J. (2016a). *How do 7th graders who take the ACT test score as 11th or 12th graders?* Iowa City, IA: ACT.

Allen, J. (2016b). *Subgroup representation and validity in ACT College Readiness Benchmark development samples*. Iowa City, IA: ACT.

Allen, J., Bolender, B., Fang, Y., Li, D., & Thompson, T. (2016). *Relating the ACT indicator Understanding Complex Texts to college course grades*. Iowa City, IA: ACT.

Allen, J., LeFebvre, M., & Mattern, K. (2016). *Predicted attainment of ACT National Career Readiness Certificates for the 2016 ACT-tested high school cohort*. Iowa City, IA: ACT.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

Cohen, J. (1968). Weighted kappa: Nominal scale agreement with provision for scaled disagreement or partial credit. *Psychological Bulletin, 70*(4), 213–220.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*(3), 297–334.

Crouse, J. D., & Allen, J. (2014). College course grades for dual enrollment students. *Community College Journal of Research and Practice, 38*(6), 494–511.

Fleiss, J. L., Levin, B., & Paik, M. C. (2013). *Statistical methods for rates and proportions* (3rd edition). Wiley & Sons. *https://www.wiley.com/en-us/Statistical+Methods+for+Rates+and+Proportions%2C+3rd+Edition-p-9781118625613*

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1–55. *https://doi.org/10.1080/10705519909540118*

Hanson, B. A. (1990). *An investigation of methods for improving estimation of test score distributions*. Iowa City, IA: ACT.

Hanson, B. A., & Brennan, R. L. (1990). An investigation of classification consistency indexes estimated under alternative strong true score models. *Journal of Educational Measurement, 27*(4), 345–359.

International Organization for Standardization. (2018). ISO/IEC 27001 family-information security management systems. Retrieved from *https://www.iso.org/isoiec-27001-information-security.html*.

Joint Committee on Testing Practices. (2004). *Code of fair testing practices in education.* Washington, DC: Joint Committee on Testing Practices. *https://www.apa.org/science/programs/testing/fair-testing.pdf.*

Kolen, M. J. (1984). Effectiveness of analytic smoothing in equipercentile equating. *Journal of Educational Statistics, 9*(1), 25–44.

Kolen, M. J. (1988). Defining score scales in relation to measurement error. *Journal of Educational Measurement, 25*(2), 97–110.

Kolen, M. J. (1991). Smoothing methods for estimating test score distributions. *Journal of Educational Measurement, 28*(3), 257–282.

Kolen, M.J. & Brennan, R.L. (2014). *Test equating, scaling, and linking: Methods and practices* (3rd ed.). New York, NY: Springer.

Kolen, M. J., & Hanson, B. A. (1989). Scaling the ACT assessment. In R. L. Brennan (Ed.), *Methodology used in scaling the ACT Assessment and P-ACT+* (pp. 35–55). Iowa City, IA: ACT.

Kolen, M. J., Hanson, B. A., & Brennan, R. L. (1992). Conditional standard errors of measurement for scale scores. *Journal of Educational Measurement, 29*(4), 285–307.

Li, D., Yi, Q., & Harris, D. (2017). *Evidence for paper and online ACT® comparability: Spring 2014 and 2015 mode comparability studies*. Iowa City, IA: ACT.

Livingston, S. A., & Lewis, C. (1995). Estimating the consistency and accuracy of classifications based on test scores. *Journal of Educational Measurement, 32*(2), 179–197.

Lord, F. M. (1965). A strong true score theory, with applications. *Psychometrika, 30*(3), 239–270.

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, PA: Addison-Wesley.

Millman, J., & Greene, J. (1989). The specification and development of tests of achievement and ability. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 335–366). New York, NY: American Council on Education, and Macmillan Publishing Company.

Moore, J. L., & Cruce, T. (2017). *Does opting into a search service provide benefits to students?* Iowa City, IA: ACT.

National Institute of Standards and Technology. (2017). Information technology laboratory: Computer security resource center. Retrieved from http://csrc.nist.gov/publications/PubsSPs.html.

NCME Ad Hoc Committee on the Development of a Code of Ethics. (1995). *Code of professional responsibilities in educational measurement.* Washington, DC: National Council on Measurement in Education.

Powers, S., Li, D., Suh, H., & Harris, D. J. (2016). *ACT reporting category interpretation guide: Version 1.0*. Iowa City, IA: ACT.

Radunzel, J., & Fang, Y. (2018). *Updating the Progress Toward the ACT National Career Readiness Certificate Indicator*. Iowa City, IA: ACT.

Radunzel, J., Mattern, K., Crouse, J., & Westrick, P. (2015). *Development and validation of a STEM Benchmark based on the ACT STEM score*. Iowa City, IA: ACT.

Radunzel, J., Westrick, P., Bassiri, D., & Li, D. (2017). *Development and validation of a preliminary ELA readiness benchmark based on the ACT ELA score.* Iowa City, IA: ACT.

Sawyer, R. L. (1989a). Research design and sampling for the Academic Skills Study. In R. L. Brennan (Ed.), *Methodology used in scaling the ACT assessment and P-ACT+* (pp. 19–33). Iowa City, IA: ACT.

Sawyer, R. (1989b). *Validating the use of ACT Assessment score and high school grades for remedial course placement in college*. Iowa City, IA: ACT.

Steedle, J., Pashley, P., & Cho, YoungWoo. (2020). *Three studies of comparability between paper-based and computer-based testing for the ACT*. Iowa City, IA: ACT. *https://www.act.org/content/dam/act/unsecured/documents/R1847-three-comparability-studies-2020-12.pdf*.

Wai, J. & Allen, J. (2019). What boosts talent development? Examining predictors of academic growth in secondary school among academically advanced youth across 21 years. *Gifted Child Quarterly, 63*(4).

Wang, M. W., & Stanley, J. L. (1970). Differential weighting: A review of methods and empirical studies. *Review of Educational Research, 40*(5), 663–705.

Webb, N. L. (2002). *Depth-of-knowledge levels for four content areas* [Unpublished manuscript]. Retrieved from: *http://ossucurr.pbworks.com/w/file/fetch/49691156/Norm%20web%20dok%20by%20subject%20area.pdf*.

Westrick, P. A. (2017). Reliability estimates for undergraduate GPA. *Educational Assessment*, *22*(4), 231–252.

Westrick, P. A., Le, H., Robbins, S. B., Radunzel, J. M., & Schmidt, F. L. (2015). College performance and retention: A meta-analysis of the predictive validities of ACT scores, high school grades, and SES. *Educational Assessment*, *20*(1), 23–45.

Appendix A

Wisconsin State Supplement: 2024–2025 Achievement Summaries and Test Characteristics

Student Participation

WDPI has contracted with ACT to provide all Wisconsin 11th grade students an opportunity to participate in a school-day administration of the ACT test. Wisconsin students are required to participate as part of the graduation requirements in Wisconsin. Historically, ACT has advised students to take the ACT after they have completed a substantial portion of the coursework covered by its tests. Given the curriculum of most secondary schools and the course of study followed by most students, this point is usually reached by spring of the junior year.

Self-reported data describing the ACT examinee population for the 2025 Wisconsin junior class are presented in Table 1. A list and count of students' approved accommodations are provided in Table 2. These data are based on the 63,361 Wisconsin 11th grade students earning a Composite score who participated in the state-sponsored spring administration of the ACT.

Table 1. Demographic Characteristics of Wisconsin State Contract Spring 2025 ACT Testers

Demographic	%[a]	N
Gender		
Female	47	29,688
Male	49	31,143
Other Gender	1	388
No response	1	905
Prefer not to respond	2	1,237
Grade level when tested		
Junior	100	63,361
Racial-ethnic background		
African American/Black	7	4,379
White	64	40,262
American Indian/Alaska Native	1	525
Hispanic/Latino	14	8,702
Asian	4	2,486
Native Hawaiian/Other Pacific Islander	< 1	49
Two or more races	5	3,164
Prefer no response/blank	6	3,794

[a] Due to rounding, some columns may not add to exactly 100%.
*Information in this table can also be found in *Table 1.1*.

Table 2. List of Approved Accommodations for the 2025 ACT-Tested High School Junior Class

Accommodation Description	N
ASL sign language interpreter for the entire test	9
Assistive device/technology (furniture, AAC, adaptive keyboard or mouse, switch, etc.)	15
Bilingual word-to-word dictionary (ACT-authorized)	879
Braille with tactile graphics (UEB with Nemeth), entire test	9
Breaks as needed/stop-the-clock, standard time (requires single room)	291
Color contrast/color overlay	13
Computer for writing constructed response (paper-based testing)	121
Double time on the writing section only (single day)	12
Double time, multiple days	540
Food, drink, medication for examinees with medical needs	172
Hearing assistive devices (audio amplification, FM/DM system)	38
Human reader who reads the entire test	94
Large Print	24
Mark answers in test booklet (no scantron)	5
Medical monitoring device (not physically attached to the examinee - requires single room)	120
Noise buffers/ear plugs (requires single room)	39
One and one-half time, multiple days	802
One and one-half time, single day	5,082
One-to-one testing	441
OTHER	204
Permission to stand during testing (requires single room)	21
Pre-recorded audio	709
Preferential seating	415
Printed copy of verbal instructions	15
Read aloud to self	12
Screen reader software for low vision or blindness (Braille/tactile graphic supplement required)	2
Scribe	62
Sign language interpreter, for verbal instructions	8
Small group testing	6,557
Speech-to-text dictation software	156
Standard time, multiple days	65
Standard time, single day	102

Accommodation Description	N
Test location - Administration at home or care facility	5
Text-to-speech/Pre-recorded Audio	1,977
Time remaining indicator (countdown timer, notecard with time remaining, tap on shoulder, etc.)	16
Translated test directions (Arabic)	10
Translated test directions (Chinese Cantonese Simp)	1
Translated test directions (Chinese Mandarin Simp)	8
Translated test directions (Chinese Mandarin Trad)	3
Translated test directions (French)	6
Translated test directions (German)	1
Translated test directions (Haitian Creole)	3
Translated test directions (Hmong Daw (White))	7
Translated test directions (Japanese)	2
Translated test directions (Portuguese)	3
Translated test directions (Russian)	4
Translated test directions (Somali)	3
Translated test directions (Spanish)	619
Translated test directions (Tagalog)	1
Translated test directions (Vietnamese)	11
Triple time, multiple days	2,833
Visual Environment	10
Wheelchair accessibility	13

Student Performance

Summary statistics in Table 3 provide an overview of student performance at the overall levels by standard test form as well as the accommodated test form. Student self-reported demographics are used to show summaries by gender, race/ethnicity, and English language learner status. *Chapter 5* provides additional information regarding scoring and reporting including an overview of the score reports and college readiness benchmarks.

Table 3. Summary Statistics of the ACT Test Score Distributions for the Wisconsin Spring 2025 Administration

Standard Forms (as referenced in *Table 5.3*)								
Statistic	English	Mathematics	Reading	Science	Composite	Writing	ELA	STEM
N	59,489	59,489	59,489	59,489	59,489	58,220	58,220	59,489
Mean	18.55	19.37	19.71	19.98	19.53	6.66	18.99	19.93
SD	6.48	5.68	6.45	5.52	5.50	1.83	5.69	5.31
Skewness	0.49	0.75	0.41	0.43	0.56	-0.41	0.17	0.66
Kurtosis	-0.27	0.09	-0.54	0.30	-0.25	-0.00	-0.44	0.11
Accommodated Forms (as referenced in *Table 5.4*)								
Statistic	English	Mathematics	Reading	Science	Composite	Writing	ELA	STEM
N	3,872	3,872	3,872	3,872	3,872	3,583	3,583	3,872
Mean	13.31	15.26	14.94	15.88	14.97	4.47	13.06	15.82
SD	4.75	3.73	5.78	4.70	4.20	1.84	5.07	3.89
Skewness	1.71	2.01	1.45	1.22	1.93	0.39	1.24	1.87
Kurtosis	3.54	6.33	2.14	2.43	4.23	-0.44	1.56	4.80
English Language Learners								
Statistic	English	Mathematics	Reading	Science	Composite	Writing	ELA	STEM
N	2,392	2,392	2,392	2,392	2,392	2,252	2,252	2,392
Mean	14.65	16.58	15.70	17.10	16.13	5.74	15.33	17.09
SD	4.99	4.53	4.96	4.48	4.15	1.97	4.92	4.15
Skewness	0.86	1.23	0.89	0.51	1.03	-0.18	0.40	1.04
Kurtosis	0.57	1.97	0.54	0.43	0.81	-0.59	-0.32	1.17
Standard Forms: Male								
Statistic	English	Mathematics	Reading	Science	Composite	Writing	ELA	STEM
N	28,956	28,956	28,956	28,956	28,956	28,101	28,101	28,956
Mean	17.88	19.81	19.24	20.10	19.38	6.34	18.25	20.21
SD	6.39	5.99	6.54	5.82	5.67	1.89	5.77	5.62
Skewness	0.57	0.70	0.47	0.45	0.60	-0.30	0.25	0.64
Kurtosis	-0.17	-0.13	-0.56	0.13	-0.28	-0.22	-0.47	-0.08
Standard Forms: Female								
Statistic	English	Mathematics	Reading	Science	Composite	Writing	ELA	STEM
N	28,203	28,203	28,203	28,203	28,203	27,844	27,844	28,203
Mean	19.15	18.93	20.10	19.84	19.62	6.99	19.70	19.63
SD	6.46	5.29	6.26	5.16	5.28	1.70	5.49	4.93
Skewness	0.42	0.75	0.38	0.36	0.53	-0.47	0.14	0.64
Kurtosis	-0.30	0.23	-0.48	0.44	-0.20	0.32	-0.36	0.24

Standard Forms: American Indian/Alaska Native								
Statistic	English	Mathematics	Reading	Science	Composite	Writing	ELA	STEM
N	469	469	469	469	469	447	447	469
Mean	15.11	16.29	16.62	17.00	16.38	5.83	15.81	16.90
SD	5.10	4.03	5.24	4.34	4.07	1.87	4.83	3.82
Skewness	0.76	0.98	0.67	0.44	0.99	-0.14	0.47	0.90
Kurtosis	0.76	1.46	0.22	1.05	1.03	-0.38	0.03	1.27
Standard Forms: Asian								
Statistic	English	Mathematics	Reading	Science	Composite	Writing	ELA	STEM
N	2,361	2,361	2,361	2,361	2,361	2,329	2,329	2,361
Mean	19.15	20.40	19.62	20.34	20.01	6.94	19.44	20.63
SD	7.16	6.58	6.70	5.92	6.12	1.86	6.03	5.96
Skewness	0.62	0.77	0.56	0.63	0.74	-0.30	0.36	0.78
Kurtosis	-0.34	-0.29	-0.45	0.28	-0.21	0.03	-0.32	-0.05
Standard Forms: Black/African American								
Statistic	English	Mathematics	Reading	Science	Composite	Writing	ELA	STEM
N	4,085	4,085	4,085	4,085	4,085	3,920	3,920	4,085
Mean	13.89	15.06	15.47	15.97	15.22	5.50	14.63	15.76
SD	4.87	3.61	4.93	4.15	3.77	1.86	4.67	3.48
Skewness	0.97	1.08	0.96	0.37	1.21	-0.04	0.64	1.04
Kurtosis	1.48	4.41	1.17	1.53	2.40	-0.55	0.36	3.31
Standard Forms: Hispanic/Latino								
Statistic	English	Mathematics	Reading	Science	Composite	Writing	ELA	STEM
N	7,981	7,981	7,981	7,981	7,981	7,688	7,688	7,981
Mean	15.85	16.85	17.30	17.72	17.06	6.22	16.78	17.54
SD	5.62	4.50	4.70	4.76	4.60	1.87	5.21	4.31
Skewness	0.86	1.17	0.74	0.55	0.96	-0.35	0.41	1.03
Kurtosis	0.59	1.76	0.11	0.92	0.79	-0.22	-0.08	1.43
Standard Forms: Native Hawaiian/Other Pacific Islander								
Statistic	English	Mathematics	Reading	Science	Composite	Writing	ELA	STEM
N	46	46	46	46	46	44	44	46
Mean	16.63	16.98	16.70	17.98	17.28	6.34	17.09	17.72
SD	5.87	5.67	5.48	6.14	5.25	1.68	4.95	5.58
Skewness	1.14	0.86	0.99	0.63	1.30	-0.94	0.69	0.99
Kurtosis	1.11	2.91	1.02	0.82	1.93	-0.01	0.19	2.27
Standard Forms: White								
Statistic	English	Mathematics	Reading	Science	Composite	Writing	ELA	STEM
N	38,153	38,153	38,153	38,153	38,153	37,628	37,628	38,153
Mean	19.79	20.52	20.86	21.06	20.68	6.92	20.09	21.04
SD	6.32	5.62	6.35	5.38	5.35	1.72	5.44	5.20
Skewness	0.36	0.61	0.27	0.36	0.42	-0.45	0.09	0.54
Kurtosis	-0.35	-0.21	-0.63	0.27	-0.37	0.26	-0.40	-0.06

Standard Forms: Two or more races								
Statistic	English	Mathematics	Reading	Science	Composite	Writing	ELA	STEM
N	3,016	3,016	3,016	3,016	3,016	2,961	2,961	3,016
Mean	18.40	18.77	19.87	19.67	19.30	6.63	18.94	19.47
SD	6.52	5.48	6.49	5.55	5.47	1.87	5.73	5.22
Skewness	0.51	0.93	0.42	0.48	0.63	-0.35	0.23	0.80
Kurtosis	-0.25	0.52	-0.54	0.37	-0.12	-0.06	-0.48	0.42
Accommodated Forms: Male								
Statistic	English	Mathematics	Reading	Science	Composite	Writing	ELA	STEM
N	2,187	2,187	2,187	2,187	2,187	1,973	1,973	2,187
Mean	12.84	15.18	14.32	15.64	14.63	4.07	12.17	15.66
SD	4.35	3.61	5.37	4.53	3.91	1.71	4.61	3.74
Skewness	1.96	2.08	1.62	1.27	2.17	0.59	1.47	1.98
Kurtosis	5.29	6.58	3.12	2.85	5.75	-0.13	2.78	5.44
Accommodated Forms: Female								
Statistic	English	Mathematics	Reading	Science	Composite	Writing	ELA	STEM
N	1,485	1,485	1,485	1,485	1,485	1,428	1,428	1,485
Mean	13.81	15.28	15.58	16.09	15.30	5.01	14.11	15.95
SD	4.96	3.76	5.93	4.75	4.31	1.85	5.25	3.93
Skewness	1.42	1.87	1.25	1.10	1.68	0.11	0.97	1.72
Kurtosis	2.19	5.92	1.36	1.97	3.05	-0.05	0.67	4.18
Accommodated Forms: American Indian/Alaska Native								
Statistic	English	Mathematics	Reading	Science	Composite	Writing	ELA	STEM
N	56	56	56	56	56	51	51	56
Mean	12.21	14.00	13.41	14.14	13.59	4.16	11.73	14.29
Std Dev	2.95	1.71	3.62	2.97	2.04	1.69	3.16	1.89
Skewness	1.11	-0.39	1.10	-0.31	1.13	0.26	0.82	0.08
Kurtosis	2.17	1.78	2.03	-0.04	2.35	-0.70	1.26	-0.01
Accommodated Forms: Asian								
Statistic	English	Mathematics	Reading	Science	Composite	Writing	ELA	STEM
N	125	125	125	125	125	119	119	125
Mean	13.31	15.35	13.93	15.34	14.56	4.75	12.89	15.62
Std Dev	3.92	4.41	4.72	4.02	3.60	1.62	4.26	3.73
Skewness	1.24	2.14	1.17	0.74	1.72	0.11	0.88	1.98
Kurtosis	2.01	6.19	1.39	0.95	3.26	-0.60	0.50	5.36
Accommodated Forms: Black/African American								
Statistic	English	Mathematics	Reading	Science	Composite	Writing	ELA	STEM
N	294	294	294	294	294	269	269	294
Mean	11.48	14.00	12.64	14.16	13.19	3.77	10.73	14.33
Std Dev	2.63	2.40	3.66	3.11	2.20	1.49	3.23	2.27
Skewness	0.86	0.64	1.30	0.65	1.88	0.63	0.92	1.35
Kurtosis	1.65	7.61	2.57	1.52	6.07	-0.03	1.05	4.87

Accommodated Forms: Hispanic/Latino								
Statistic	English	Mathematics	Reading	Science	Composite	Writing	ELA	STEM
N	721	721	721	721	721	654	654	721
Mean	12.48	14.62	14.14	15.08	14.21	4.54	12.66	15.12
Std Dev	3.66	2.61	4.69	3.87	3.07	1.68	4.20	2.86
Skewness	1.61	1.98	1.46	0.90	2.04	0.07	1.00	1.83
Kurtosis	4.27	10.51	3.29	2.21	6.61	-0.80	1.32	7.35
Accommodated Forms: Native Hawaiian/Other Pacific Islander								
Statistic	English	Mathematics	Reading	Science	Composite	Writing	ELA	STEM
N	3	3	3	3	3	2	2	3
Mean	15.67	15.67	16.33	18.00	16.33	6.50	19.50	17.00
Std Dev	6.81	2.31	7.64	6.24	5.69	0.71	3.54	4.36
Skewness	-1.36	-1.73	-0.94	-1.29	-1.21	.	.	-1.63
Kurtosis	.	.	.	.	.	.	.	.
Accommodated Forms: White								
Statistic	English	Mathematics	Reading	Science	Composite	Writing	ELA	STEM
N	2,109	2,109	2,109	2,109	2,109	1,986	1,986	2,109
Mean	13.89	15.69	15.69	16.49	15.57	4.59	13.63	16.34
SD	5.27	4.08	6.31	5.07	4.67	1.92	5.49	4.28
Skewness	1.54	1.75	1.30	1.14	1.68	0.41	1.16	1.65
Kurtosis	2.47	4.39	1.30	1.85	2.80	-0.43	1.06	3.41
Accommodated Forms: Two or more races								
Statistic	English	Mathematics	Reading	Science	Composite	Writing	ELA	STEM
N	148	148	148	148	148	135	135	148
Mean	13.91	15.47	15.31	15.95	15.24	4.42	13.35	15.96
SD	4.81	3.58	5.82	4.78	4.21	1.77	5.11	3.84
Skewness	1.43	1.53	1.25	1.18	1.79	0.39	1.17	1.72
Kurtosis	2.94	2.78	1.43	2.18	3.43	-0.41	1.39	3.14

Effective Weights and Correlations

Table 4 presents the ACT scale score covariances from the primary standard test form administered in Wisconsin during the 2024–2025 academic year. Table 5 shows effective weights for the Composite and ELA scores based on the primary standard test form administered in Wisconsin during the 2024–2025 academic year. Additional information including regarding the uses of covariances, effective weights, and correlations can be found in *Section 5.2.6* of this technical manual. The subject-level correlations in Table 6 can also be found in *Chapter 5* while the detailed reporting category correlations in Table 6 are presented only in this appendix.

Table 4. Scale Score Covariances for Multiple-Choice Tests from the Primary ACT Test Form

Test	English	Mathematics	Reading	Science
English	42.57	28.92	35.20	28.79
Mathematics	28.92	33.89	26.83	25.78
Reading	35.20	26.83	40.71	28.32
Science	28.79	25.78	28.32	30.99

Note: Information in this table can also be found in *Table 5.5*.

Table 5. Effective Weights of the ACT Tests from the Primary Test Form

Test	Composite	ELA
English	0.27	0.35
Mathematics	0.23	--
Reading	0.26	0.33
Science	0.23	--
Writing	--	0.32

Note: Information in this table can also be found in *Table 5.6*.

Table 6. Correlations Among the ACT Test Scores Administered in Wisconsin in Spring 2025 (N=58,220)

Score	English	Mathematics	Reading	Science	Composite	Writing	ELA
English	1.00	0.75	0.82	0.77	0.93	0.56	0.91
Mathematics		1.00	0.70	0.79	0.88	0.49	0.74
Reading			1.00	0.77	0.91	0.53	0.89
Science				1.00	0.91	0.51	0.78
Composite					1.00	0.57	0.92
Writing						1.00	0.81
ELA							1.00

English

Reporting categories	PoW	KLA	CoE
PoW	1.00	0.69	0.78
KLA		1.00	0.70
CoE			1.00

Mathematics

Reporting categories	PHM	NAQ	Algebra	Functions	Geometry	SAP	IES	Modeling
PHM	1.00	0.75	0.83	0.79	0.74	0.74	0.81	0.81
NAQ		1.00	0.55	0.49	0.45	0.46	0.59	0.56
Algebra			1.00	0.55	0.50	0.54	0.69	0.63
Functions				1.00	0.48	0.48	0.61	0.61
Geometry					1.00	0.41	0.60	0.68
SAP						1.00	0.63	0.65
IES							1.00	0.83
Modeling								1.00

Reading

Reporting categories	KID	CAS	IOK
KID	1.00	0.72	0.57
CAS		1.00	0.50
IOK			1.00

Science

Reporting categories	IOD	SIN	EMI
IOD	1.00	0.63	0.56
SIN		1.00	0.58
EMI			1.00

Note: Information in this table can also be found in *Table 5.7*.

Writing Domain Scores

In addition to the overall writing test score, scores are also reported for four domains: Ideas & Analysis, Development & Support, Organization, and Language Use & Conventions. These domains measure essential skills and abilities that are required for college and career success.

Table 7 presents the summary statistics of writing domain scores and the overall writing scores based on all test forms administered in Wisconsin during the 2024–2025 academic year. Table 8 presents the correlations among these scores. Additional information regarding the Writing scores can be found in *Chapter 5*.

Table 7. Summary Statistics of the ACT Writing and Writing Domain Score Distributions for the Wisconsin Spring 2025 Administration

Statistic	Ideas & Analysis	Development & Support	Organization	Language Use & Conventions	Writing score
N	58,220	58,220	58,220	58,220	58,220
Mean	6.63	6.24	6.52	6.88	6.66
SD	1.91	1.89	1.87	1.74	1.83
Skewness	-0.44	-0.25	-0.48	-0.43	-0.41
Kurtosis	-0.03	-0.36	-0.08	0.30	-0.00

Note: Information in this table can also be found in Table 5.8.

Table 8. Correlations Among the ACT Writing and Writing Domain Scores for the Wisconsin Spring 2025 Administration

Score	Ideas & Analysis	Development & Support	Organization	Language Use & Conventions	Writing score
Ideas & Analysis	1.00	0.95	0.98	0.95	0.99
Development & Support		1.00	0.95	0.91	0.95
Organization			1.00	0.94	0.98
Language Use & Conventions				1.00	0.97
Writing score					1.00

Note: Information in this table can also be found in *Table 5.9*.

Internal Structure

To evaluate the internal structure of the ACT test, ACT estimated confirmatory factor analysis (CFA) models for the main form. CFA tests whether a theoretically driven model of internal structure is consistent with observed data for a test or measure. All models were fit using operationally scored items and a weighted least-squares estimator. These models are unidimensional models, where all items load on a single latent factor representing knowledge or ability in a subject. For example, the structure model for the science test on the primary standard form for Wisconsin is depicted in Figure 1. The models test the unidimensionality assumption. Table 9 reports the model fit statistics of these models including chi-square, root mean square error of approximation (RMSEA), and comparative fit index (CFI). Hu and Bentler's (1999) guidelines for acceptable fit include a non-significant χ^2, RMSEA values of < 0.06, and CFI values of > 0.95. The results of χ^2 tests are known to be influenced by sample size, so even models with good fit can have statistically significant results when the sample size is large. Based on the RMSEA, the model fit is acceptable.

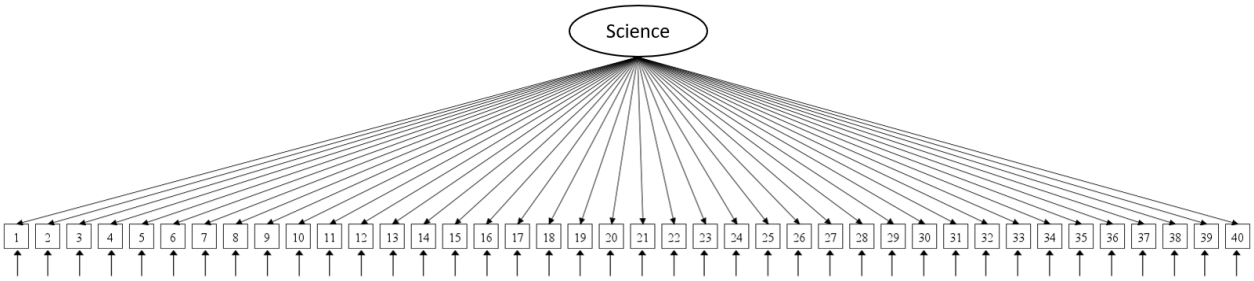


Figure 1. Latent Factor Model for the ACT Science Test on the Primary Form for Wisconsin

Table 9. Fit Statistics of Unidimensional Models of the ACT Tests

	Chi square	DF	*P*-value of Chi-square	RMSEA	CFI
English	45914.229	2700	0	0.03	0.91
Mathematics	24215.132	1710	0	0.03	0.93
Reading	8743.610	740	0	0.03	0.96
Science	9230.930	740	0	0.03	0.96

To test the hypothesis that the subscore reporting structure is consistent with the internal structure shown in observed data, ACT conducted a series of confirmatory factory analyses. For each ACT test section, internal structure models were built based on the test blueprint and reporting category classification. These models tested each reporting category separately. For example, the science test reporting category models for the primary standard form for Wisconsin are depicted in Figures 2–4. Each of these latent factors is measured by a certain number of observed variables, which are test items. Each item measures only one latent factor, and each latent factor was fit separately because the three latent factors can be highly correlated with each other.

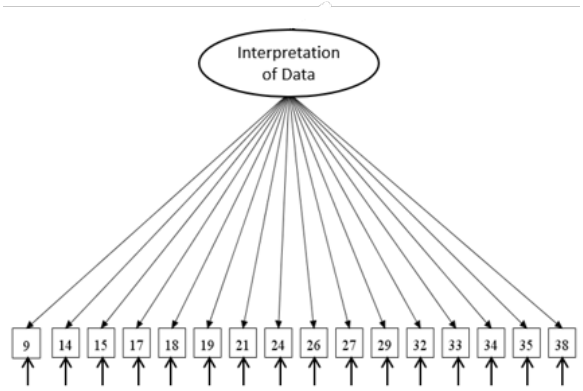


Figure 2. Latent Factor Model for the ACT Interpretation of Data Reporting Category for the Primary Form for Wisconsin

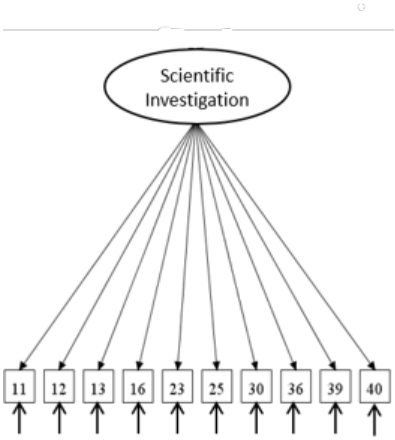


Figure 3. Latent Factor Model for the ACT Scientific Investigation Reporting Category for the Primary Form for Wisconsin

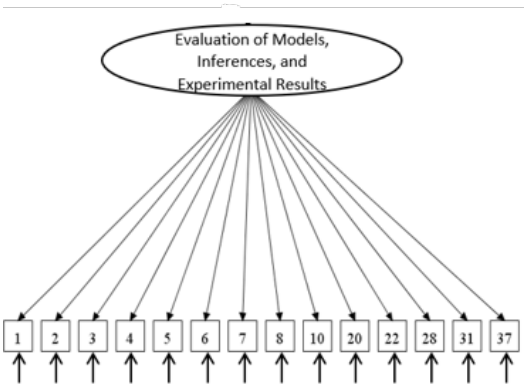


Figure 4. Latent Factor Model for the ACT Evaluation of Models, Inferences, and Experimental Results Reporting Category for the Primary Form for Wisconsin

The model fit for the reporting category models can be seen in Table 10. Most of the reporting category models also fit well based on Hu and Bentler's (1999) guidelines for RMSEA and CFI. The average standardized loadings for the reporting category models in Table 11 ranged from 0.39 to 0.61.

Table 10. Model Fit Statistics of Independent Reporting Category Models

Test section	Reporting category	χ^2	*df*	*p*-value of χ^2 test	RMSEA	CFI
English	Production of Writing	3101.460	209	0	0.03	0.97
	Knowledge of Language	3118.577	54	0	0.06	0.94
	Conventions of Standard English	11548.138	779	0	0.03	0.94
Mathematics	Number & Quantity	137.945	9	0	0.03	0.98
	Algebra	205.578	20	0	0.03	0.99
	Functions	417.960	20	0	0.04	0.93
	Geometry	478.543	20	0	0.04	0.92
	Statistics & Probability	95.663	9	0	0.03	0.99
	Integrating Essential Skills	3767.527	252	0	0.03	0.97
Reading	Key Ideas & Details	1968.549	189	0	0.03	0.98
	Craft & Structure	912.809	54	0	0.03	0.98
	Integration of Knowledge and Ideas	108.692	14	0	0.02	0.99
Science	Interpretation of Data	1333.016	104	0	0.03	0.96
	Scientific Investigation	290.627	35	0	0.02	0.98
	Evaluation of Models, Inferences & Experimental Results	893.140	77	0	0.03	0.98

Table 11. Average Factor Loadings from Independent Reporting Category Models

English						
Reporting categories	Production of Writing		Knowledge of Language		Conventions of Standard English	
Average factor loadings	0.54		0.61		0.51	
Mathematics						
Reporting categories	Number & Quantity	Algebra	Functions	Geometry	Statistics & Probability	Integrating Essential Skills
Average factor loadings	0.49	0.57	0.40	0.39	0.54	0.52
Reading						
Reporting categories	Key Ideas & Details		Craft & Structure		Integration of Knowledge & Ideas	
Average factor loadings	0.53		0.55		0.47	

Science			
Reporting categories	Interpretation of Data	Scientific Investigation	Evaluation of Models, Inferences & Experimental Results
Average factor loadings	0.46	0.45	0.52

Operational DIF Analyses

Items are analyzed and reviewed for DIF after field testing and each operational administration. Table 12 provides the DIF analysis results based on the Wisconsin student data from the spring 2025 administration. Additional information regarding DIF analyses can be found in *Chapter 2*.

Table 12. ACT Test Items Exhibiting DIF based on 2025 Wisconsin Student Data

Test	Reference group	Focal group	Number of Items	No DIF	DIF
English	Male	Female	75	75	0
	Non-EL	English learner (EL)	75	75	0
	White	African-American	75	75	0
	White	Asian	75	75	0
	White	Hispanic	75	75	0
	White	Two or more races	75	75	0
Mathematics	Male	Female	60	60	0
	Non-EL	English learner	60	60	0
	White	African-American	60	60	0
	White	Asian	60	60	0
	White	Hispanic	60	60	0
	White	Two or more races	60	60	0
Reading	Male	Female	40	40	0
	Non-EL	English learner	40	40	0
	White	African-American	40	40	0
	White	Asian	40	40	0
	White	Hispanic	40	40	0
	White	Two or more races	40	40	0
Science	Male	Female	40	40	0
	Non-EL	English learner	40	40	0
	White	African-American	40	40	0
	White	Asian	40	40	0
	White	Hispanic	40	40	0
	White	Two or more races	40	40	0

Note: Information in this table can also be found in *Table 2.3*.

Reliability and SEM for the ACT Test Scores

Scale score reliability estimates and SEM for the four ACT multiple-choice tests (English, mathematics, reading, and science), as well as for the Composite, and ELA scores are provided in Table 13, and they are also provided by subgroups in Table 14. These values were calculated based on operational test data

from the primary standard and accommodated test forms administered in the 2024-2025 academic year. See *Chapter 6* for additional information regarding reliability and measurement error.

Table 13. Scale Score Reliability and SEM for the Wisconsin Spring 2025 ACT Test Scores

		Standard		Accommodated	
Test	Number of items	Reliability	SEM	Reliability	SEM
English	75	0.93	1.70	0.79	1.76
Mathematics	60	0.91	1.77	0.74	1.70
Reading	40	0.89	2.15	0.78	2.26
Science	40	0.85	2.16	0.67	2.26
Composite	215	0.97	0.98	0.91	1.01
ELA	116	0.94	1.41	0.88	1.44
STEM	100	0.93	1.40	0.81	1.41

Note: Information in this table can also be found in *Table 6.1*.

Table 14. Reliability and Standard Error of Measurement on the ACT Test for the Wisconsin Spring 2025 Administration

	Standard form			Female			Male			African-American		
	N	SEM	Rel.	N	SEM	Rel.	N	SEM	Rel.	N	SEM	Rel.
ELA	13,829	1.412	0.939	6,476	1.409	0.934	6,766	1.414	0.941	707	1.399	0.908
English	13,829	1.704	0.931	6,476	1.692	0.931	6,766	1.714	0.929	707	1.766	0.866
Reading	13,829	2.147	0.885	6,476	2.140	0.880	6,766	2.148	0.889	707	2.016	0.835
Mathematics	13,829	1.765	0.907	6,476	1.766	0.893	6,766	1.765	0.917	707	1.764	0.772
Science	13,829	2.162	0.847	6,476	2.114	0.830	6,766	2.207	0.859	707	2.273	0.702
STEM	13,829	1.396	0.933	6,476	1.377	0.923	6,766	1.413	0.939	707	1.439	0.836

	Asian			American-Indian			Hispanic			White		
	N	SEM	Rel.	N	SEM	Rel.	N	SEM	Rel.	N	SEM	Rel.
ELA	596	1.403	0.948	104	1.408	0.905	1,990	1.411	0.924	9,030	1.413	0.934
English	596	1.711	0.945	104	1.747	0.871	1,990	1.750	0.897	9,030	1.682	0.931
Reading	596	2.085	0.906	104	2.088	0.804	1,990	2.101	0.856	9,030	2.171	0.880
Mathematics	596	1.717	0.930	104	1.760	0.818	1,990	1.744	0.852	9,030	1.767	0.907
Science	596	2.174	0.859	104	2.165	0.750	1,990	2.182	0.783	9,030	2.134	0.844
STEM	596	1.385	0.945	104	1.395	0.870	1,990	1.397	0.894	9,030	1.385	0.932

	Two or more races			English learners			Accommodated form		
	N	SEM	Rel.	N	SEM	Rel.	N	SEM	Rel.
ELA	710	1.409	0.943	462	1.407	0.919	1,772	1.439	0.876
English	710	1.706	0.930	462	1.763	0.880	1,772	1.762	0.785
Reading	710	2.130	0.891	462	2.067	0.812	1,772	2.260	0.776
Mathematics	710	1.736	0.906	462	1.733	0.865	1,772	1.698	0.738
Science	710	2.188	0.850	462	2.202	0.782	1,772	2.263	0.669
STEM	710	1.397	0.931	462	1.401	0.898	1,772	1.415	0.812

Reliability and SEM for ACT Reporting Category Scores

Raw score reliability estimates, computed using coefficient alpha, and SEM were also calculated for the ACT reporting categories based on the juniors taking the primary form administered in Wisconsin in the 2024–2025 academic year. The results are listed in Table 15.

Table 15. Raw Score Reliability and SEM for the Wisconsin Spring 2025 ACT Reporting Categories

	Standard			Accommodated		
Test/reporting categories	Number of items	Reliability	SEM	Number of items	Reliability	SEM
English						
Production of Writing	22	0.82	2.01	22	0.61	2.09
Knowledge of Language	12	0.78	1.46	12	0.43	1.53
Conventions of Standard English	41	0.88	2.78	41	0.62	2.78
Mathematics						
Preparing for Higher Math	36	0.85	2.57	36	0.56	2.57
Number & Quantity	6	0.52	1.09	6	0.25	1.02
Algebra	8	0.66	1.18	8	0.20	1.17
Functions	8	0.47	1.23	8	0.16	1.26
Geometry	8	0.43	1.17	8	0.12	1.28
Statistics & Probability	6	0.56	1.04	6	0.36	0.98
Integrating Essential Skills	24	0.83	2.11	24	0.65	1.99
Modeling	25	0.78	2.17	23	0.60	2.05
Reading						
Key Ideas & Details	21	0.81	1.99	22	0.69	2.06
Craft & Structure	12	0.73	1.49	11	0.47	1.48
Integration of Knowledge & Ideas	7	0.52	1.17	7	0.34	1.13
Science						
Interpretation of Data	16	0.71	1.78	20	0.64	1.96
Scientific Investigation	10	0.59	1.42	10	0.27	1.44
Evaluation of Models, Inferences & Experimental Results	14	0.73	1.62	10	0.49	1.33

Note: Information in this table can also be found in *Table 6.2*.

Agreement Indices for the ACT Writing Test Scores

Two major sources can contribute to the measurement error of a writing test score: rater variability and prompt variability. To get a reliability estimate that takes into account both sources of error, a special study is needed where students are administered multiple writing prompts and student responses are rated by multiple raters. Results from these studies are reported in Chapter 6 of the *ACT Technical Manual*. With Wisconsin data, where each student takes a single prompt, only rater agreement is reported based on data from the students taking the primary forms in the 2024–2025 academic year. Table 16 lists several rater agreement indices included the perfect agreement rate, the perfect plus adjacent agreement rate, and the quadratic weighted kappa coefficient.

Table 16. Agreement Rates for the ACT Writing Domain Scores for the Wisconsin Spring 2025 Administration

Domain	Agreement index	Value
Ideas & Analysis	Perfect Agreement	0.73
	Perfect + Adjacent Agreement	>0.99
	Quadratic Weighted Kappa	0.86
Development & Support	Perfect Agreement	0.72
	Perfect + Adjacent Agreement	>0.99
	Quadratic Weighted Kappa	0.85
Organization	Perfect Agreement	0.73
	Perfect + Adjacent Agreement	>0.99
	Quadratic Weighted Kappa	0.85
Language Use & Conventions	Perfect Agreement	0.70
	Perfect + Adjacent Agreement	0.99
	Quadratic Weighted Kappa	0.82

Note: Information in this table can also be found in *Table 6.3*.

Classification Consistency

Analyses were conducted to examine the classification consistency on differentiating students into performance levels using the examinees who took the primary test forms administered in Wisconsin in the 2024–2025 academic year. The classification consistencies were calculated using the Livingston and Lewis (1995) method. This method was selected because it can be used to calculate the classification consistency of composite scores, such as the ELA score. Table 17 provides the cut scores developed by WDPI and ACT and used in these analyses. Additional information regarding classification indices can be found in Chapter 6 in *Section 6.3.2*. Classification was based on the Wisconsin ACT Performance level cut scores that were developed in a standard setting workshop conducted in the summer of 2024. Additional information regarding the workshop can be found on the WPDI website in the *2024 Wisconsin ACT Performance Level Standard Setting Report*. Table 18 summarizes the agreements between the operational test classifications, showing the percentages of students who would be placed in the same achievement level on two equivalent administrations of the test. Classification consistencies were computed for subgroups, provided in Table 19.

Table 17. ACT Performance Level Cut Scores for Wisconsin

Test	Approaching Cut Score	Meeting Cut Score	Advanced Cut Score
ELA	14	19	26
Mathematics	16	19	27
Science	17	21	26

Table 18. Classification Consistency for the Wisconsin Spring 2025 Performance Level Cut Scores

		Classification Consistency	
Test	Number of items	Two-level	Four-level
ELA	116	0.90	0.79
Mathematics	60	0.88	0.76
Science	40	0.85	0.69

Note: Information in this table can also be found in *Table 6.4*.

Table 19. Classification Consistency for the Wisconsin Spring 2025 ACT Readiness Ranges

	Standard form (n=13,829)				**Female (n=6,476)**			
	Two Levels		Four Levels		Two Levels		Four Levels	
	Agreement	Kappa	Agreement	Kappa	Agreement	Kappa	Agreement	Kappa
ELA	0.895	0.790	0.755	0.656	0.891	0.777	0.751	0.642
Mathematics	0.880	0.757	0.660	0.532	0.872	0.740	0.653	0.516
Science	0.849	0.694	0.587	0.439	0.837	0.667	0.582	0.425
	Male (n=6,766)				**African-American (n=707)**			
	Two Levels		Four Levels		Two Levels		Four Levels	
	Agreement	Kappa	Agreement	Kappa	Agreement	Kappa	Agreement	Kappa
ELA	0.899	0.797	0.758	0.665	0.921	0.746	0.768	0.639
Mathematics	0.884	0.769	0.669	0.548	0.924	0.659	0.739	0.476
Science	0.857	0.712	0.597	0.456	0.890	0.524	0.671	0.396
	Asian (n=596)				**American-Indian (n=104)**			
	Two Levels		Four Levels		Two Levels		Four Levels	
	Agreement	Kappa	Agreement	Kappa	Agreement	Kappa	Agreement	Kappa
ELA	0.909	0.818	0.749	0.656	0.903	0.759	0.785	0.673
Mathematics	0.887	0.773	0.650	0.525	0.884	0.664	0.718	0.517
Science	0.862	0.718	0.560	0.405	0.819	0.487	0.629	0.413
	Hispanic (n=1,990)				**White (n=9,030)**			
	Two Levels		Four Levels		Two Levels		Four Levels	
	Agreement	Kappa	Agreement	Kappa	Agreement	Kappa	Agreement	Kappa
ELA	0.897	0.773	0.751	0.643	0.891	0.774	0.754	0.646
Mathematics	0.893	0.724	0.669	0.491	0.866	0.730	0.651	0.518
Science	0.863	0.635	0.618	0.424	0.836	0.671	0.570	0.416
	Two or more races (n=710)							
	Two Levels		Four Levels					
	Agreement	Kappa	Agreement	Kappa				
ELA	0.900	0.801	0.758	0.666				
Mathematics	0.892	0.773	0.654	0.518				
Science	0.862	0.713	0.606	0.457				
	English learner (n=462)				**Accommodated form (n=1,772)**			
	Two Levels		Four Levels		Two Levels		Four Levels	
	Agreement	Kappa	Agreement	Kappa	Agreement	Kappa	Agreement	Kappa
ELA	0.900	0.756	0.758	0.646	0.968	0.748	0.863	0.664
Mathematics	0.901	0.741	0.670	0.495	0.956	0.722	0.757	0.413
Science	0.874	0.644	0.627	0.426	0.944	0.697	0.728	0.399

Note: Information in this table can also be found in *Table 6.5*.