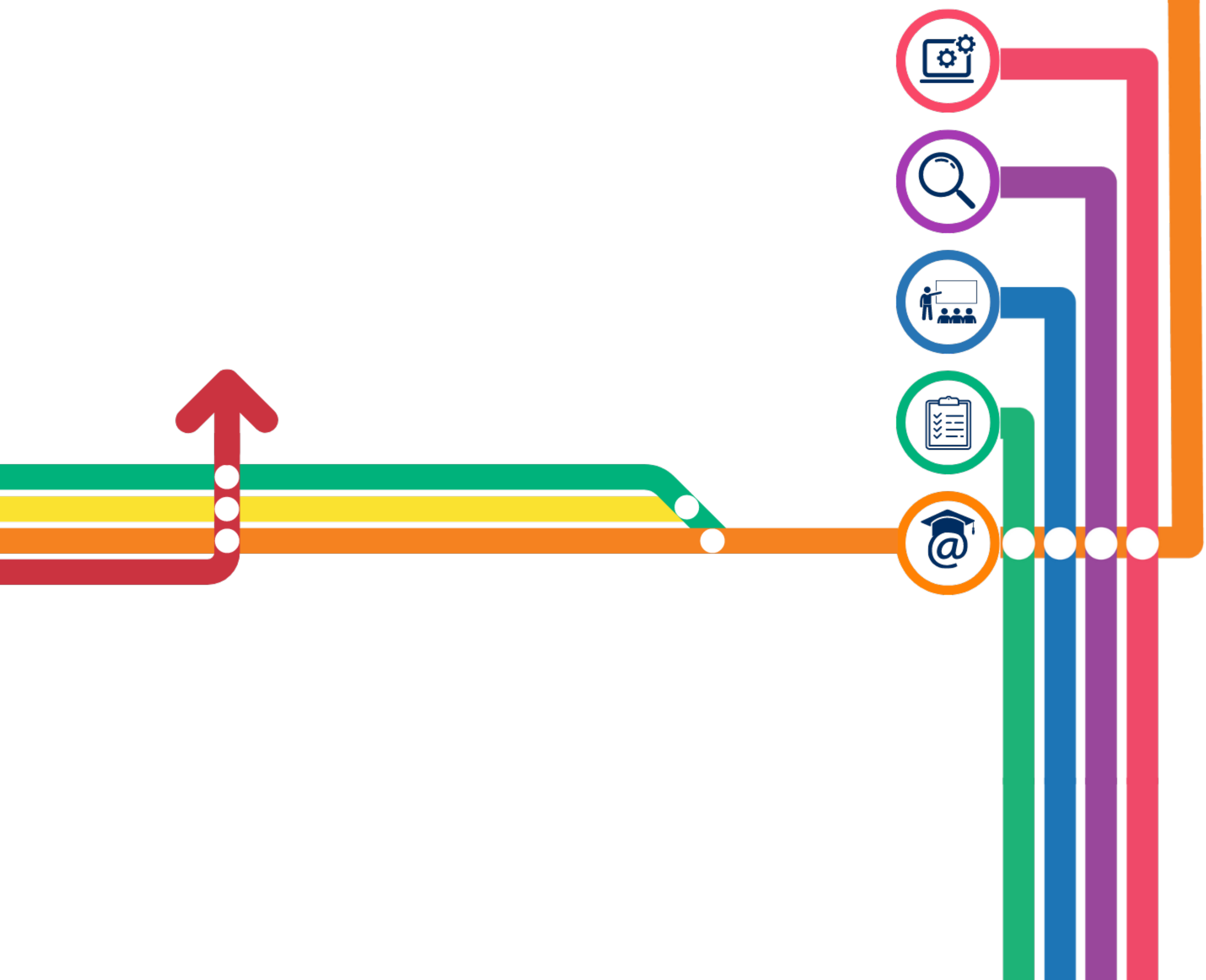
ACT® Technical Manual

February 2025



Commitment to Fair Testing

ACT endorses and is committed to complying with *The Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 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 informing 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.

Table of Contents

List of Tables

List of Figures

Preface

The principal purpose of *The ACT® Technical Manual* 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 ACT.

Please direct comments or inquiries to the address below.

Research Services
ACT, Inc.
P.O. Box 168
Iowa City, Iowa 52243-0168

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 skills (by combining mathematics and science scores), an Understanding of Complex Texts (UCT) indicator, and an ELA 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 section 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 section 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 nonacademic measures can help in the following ways:

- Students, parents, and educators can 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 can 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 can raise college awareness and exposure when all students take the ACT through state or district testing.
- Schools and districts can 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 can select students for admission who are likely to enroll at the institution and, once enrolled, likely to succeed in their courses and complete a degree at the institution.
- Postsecondary institutions can place students in first-year college courses in which they are most likely to be successful.

- Postsecondary institutions can identify students early on who are most likely to struggle academically, who may be at risk of dropping out, 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 ACT National Curriculum Survey provides empirical validation evidence related to the content of the tests. The most recent survey was released in 2020 and included responses from thousands of educators, from K–12 to college, in ELA, mathematics, reading, and science (ACT, 2020b).

The ACT National Curriculum Survey includes workforce supervisors and employees, whose responses are reviewed by SMEs and used to identify the most critical skills and knowledge required for career readiness.

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 sections tested by the ACT.

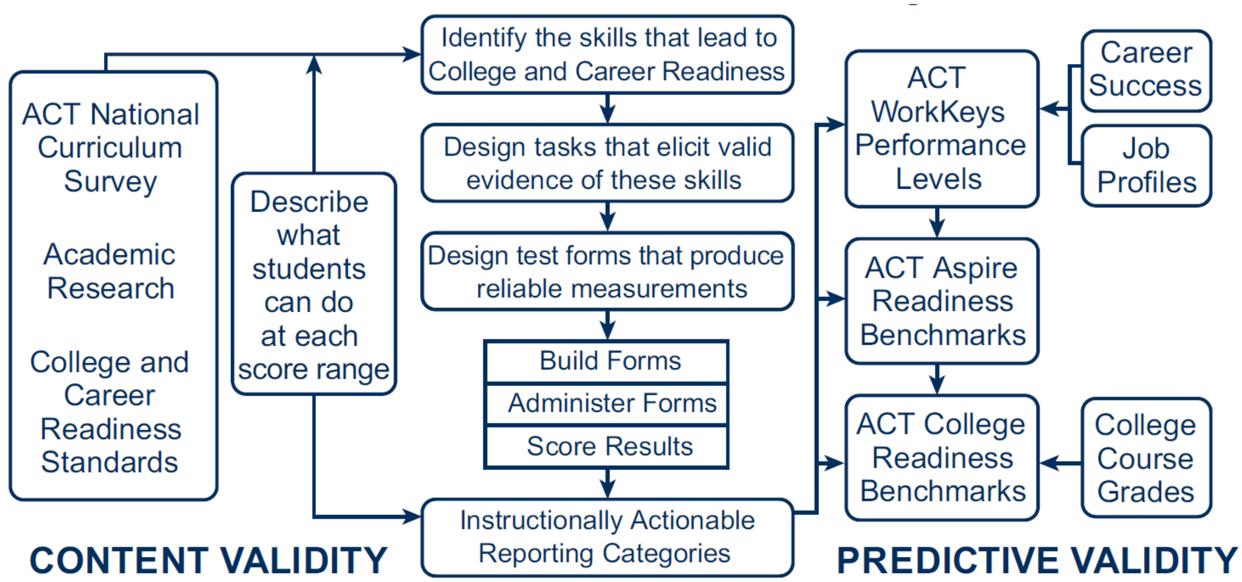
The first step in developing the ACT was to synthesize research on high-value skill targets—the skill targets that can be shown to offer the most useful evidence of college and career readiness. This evidence was obtained by organizing the knowledge and skills identified by educators and contained in 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 on 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" (p. 2). Test developers should "avoid potentially offensive content or language" (p. 4) and "evaluate the evidence to ensure that differences in performance are related to the skills being assessed" (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

Over three million students take the ACT each year. More than 3,000 postsecondary institutions (including scholarship agencies, state educational systems, individual public and private universities, four-year colleges, junior and community colleges, nursing schools, and technical schools) require or recommend that applicants submit ACT test results.

For the majority of students, postsecondary education begins shortly after high school graduation. Students typically take the ACT during their sophomore, junior, or senior year of high school or shortly before they enroll at a postsecondary institution. Thus, most students who take the ACT are between the ages of 16 and 20.

Most analyses reported in this edition of *The ACT Technical Manual* are based on approximately one million records from seven ACT forms administered from June 2023 to April 2024 to large samples during Saturday ACT National testing events. Self-reported data describing this ACT examinee sample are summarized in Table 1.1. Note that the broader ACT testing population in the United States has somewhat different demographic characteristics because it also includes students who took the ACT during the school day through ACT State and District testing.

ACT advises 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 the majority of students, this point is usually reached by spring of junior year. However, this varies from student to student and between the four academic areas measured by the ACT, so many students continue to take the ACT into senior year.

Table 1.1. Demographic Characteristics of National ACT Testers, June 2023 to April 2024

Demographic		Percentage (%)
Gender	Female	56.1
	Male	42.7
	Another gender	0.2
	Prefer not to respond	1.0
Grade level when tested	Below 9th grade	0.4
	9th grade	1.2
	10th grade	7.9
	11th grade	53.6
	12th grade	34.8
	Other	2.0
Racial/ethnic background	American Indian/Alaska Native	0.5
	Asian	7.0
	Black/African American	11.6
	Hawaii Native/other Pacific Islander	0.1
	Hispanic/Latino	12.5
	White	59.5
	Two or more races	4.3
	Prefer not to respond	4.5

Note. Due to rounding, some values may not add to exactly 100%.

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.*** This resource 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: 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 in to 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.

- ***ACT Online Prep.*** This resource provides students with an interactive test preparation experience that can be accessed anytime online and includes both structured and adaptive paths. It includes personalized learning paths, practice tests with real ACT test questions, and comprehensive content review.

- ***ACT Question of the Day.*** We post a daily test question to MyACT to provide students with an opportunity for quick daily practice. Students and teachers can opt to receive a weekly email reviewing the questions posted the week before.

- ***Powered by Kaplan.*** ACT has partnered with Kaplan to publish three official test preparation products:

    - Self-Paced Course: Delivers bite-sized video lessons on demand so students can learn anywhere, anytime.

    - Live Online Classes: Our top-rated teachers show students what to study and how to study during a series of engaging live classes.

    - Tutoring: Led by expert tutors, students learn test content and strategies in these one-on-one online tutoring sessions. Our instructors adapt to the student’s needs and provide each student with personalized attention and recommendations.

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, 2007a, 2009, 2013a, 2016b, 2020b). 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®, 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, 2020b). 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 ACT forms are constructed by selecting items from this pool; items are chosen that match the content and statistical specifications (described in Chapter 3).

Table 2.2 displays the distributions of item difficulty levels for seven forms administered from June 2023 to April 2024. 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 seven National test forms administered from June 2023 to April 2024. 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, June 2023 to April 2024

Statistic		English	Math	Reading	Science
Difficulty range	.00–.09	0.00%	0.00%	0.00%	0.00%
	.10–.19	0.19%	1.43%	0.00%	0.36%
	.20–.29	1.33%	12.14%	0.36%	4.64%
	.30–.39	4.19%	14.29%	4.29%	12.50%
	.40–.49	10.67%	14.76%	15.00%	15.00%
	.50–.59	17.14%	19.52%	17.86%	18.57%
	.60–.69	24.00%	16.67%	27.50%	20.36%
	.70–.79	27.43%	15.48%	22.86%	18.21%
	.80–.89	14.10%	5.24%	11.07%	10.36%
	.90–1.00	0.95%	0.48%	1.07%	0.00%
No. items [c]		525	420	280	280
Mean difficulty		0.65	0.53	0.64	0.58
Mean discrimination		0.44	0.44	0.42	0.42
Mean completion rate [d]		95%	95%	97%	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 underrepresentation in any category. Reviewers also look for 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 and 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 preestablished 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 percentages of ACT items that exhibited DIF according to the MH procedure for seven forms administered from June 2023 to April 2024. 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 were near or below the expected 5%. For example, in the gender DIF analysis, the percentages of items flagged ranged from 0% in science to 5.5% in mathematics. In the race/ethnicity analysis, the percentage of items flagged for Asian–White DIF sometimes exceeded the expected 5% rate (11.5% in English, 11.4% in math). These values were in the 2%–7% range in the past, so they will be monitored in coming years to determine whether this is an anomaly or a trend.

Table 2.3. ACT Test Items Exhibiting Differential Item Functioning, June 2023 to April 2024

Comparison	Favored group	English	Math	Reading	Science
Female–Male	Female	0.6%	0.5%	1.4%	—
	Male	1.0%	5.0%	0.7%	—
American Indian/Alaska Native–White	American Indian/Alaska Native	—	—	—	—
	White	—	0.5%	—	0.4%
Asian–White	Asian	6.7%	9.0%	1.1%	1.1%
	White	4.8%	2.4%	—	0.4%
Black/African American–White	Black/African American	1.1%	—	—	—
	White	4.2%	4.8%	1.4%	1.4%
Hawaii Native/other Pacific Islander–White	Hawaii Native/other Pacific Islander	0.2%	0.2%	—	0.4%
	White	0.8%	0.5%	0.4%	—
Hispanic/Latino–White	Hispanic/Latino	—	—	—	—
	White	1.0%	0.5%	0.4%	0.4%

Note. Total item counts were 525 for English, 420 for math, 280 for reading, and 280 for science.

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 a 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 Mathematics (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.

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 Progress Toward the ACT WorkKeys National Career Readiness Certificate Indicator

The Progress Toward the ACT® WorkKeys® National Career Readiness Certificate® (NCRC®) Indicator is based on students' ACT Composite scores. It provides an estimate of students' most likely performance on the three assessments that lead to the WorkKeys NCRC, 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 WorkKeys assessments: ACT® WorkKeys® Applied Math, ACT® WorkKeys® Graphic Literacy, and ACT® WorkKeys® Workplace Documents. Scores on these assessments determine an individual's certificate level—no certificate, Bronze, Silver, Gold, or Platinum. The WorkKeys NCRC gives individuals evidence that they possess the skills employers deem essential to workplace success. More information about the WorkKeys NCRC can be found at https://workforce.act.org/credential.

2.6.6 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.

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*.

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

Note. 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 8 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 8 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 & 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 8 for 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 8 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.

[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.

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.

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 ACT is administered on predetermined test dates. For the ACT National and International tests, these dates and registration deadlines are available at *www.act.org*. The ACT is administered only on those dates and times scheduled for a given test center. For ACT State and District testing, stakeholders can choose from predetermined dates and windows during the fall and spring.

4.1.2 Testing Modes

The ACT is administered on paper for National testing and online for International testing, though there are exceptions for special testing and certain accommodations. State and District testing sites have the option of administering the test on paper or online. 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

The ACT is administered at selected sites domestically and internationally. Site locations are available at *www.act.org*. Typically, the sites are K–12 public, parochial, and private schools and postsecondary institutions. To become a test center for domestic or international administrations, prospective sites must complete an establishment request, which is evaluated by ACT test administration staff, and then complete the establishment form. Each test center must undergo renewal annually.

4.1.4 Policies and Procedures

Administration Manuals

For both paper and online administrations, ACT provides test centers 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 and responsibilities of test center staff;
- prohibited behaviors;
- handling and documenting testing irregularities;
- documentation to be submitted to ACT after testing; and
- procedures for returning test materials to ACT.

Every test center staff member must read the documentation before test day and adhere to standardized procedures.

Staffing

The test coordinator is responsible for providing both the facilities and test center staff (room supervisors and proctors). In the event a center must cancel a test date to which it had committed, the test coordinator must notify ACT test administration staff immediately so ACT can secure alternate facilities and staff.

All 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 members must understand ACT policies and procedures and their own responsibilities for implementing them. It is critical that the same procedures are followed at every test center. The test coordinator is responsible for providing test center staff with the proper manuals and training prior to test day.

All 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 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 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–National and Special 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

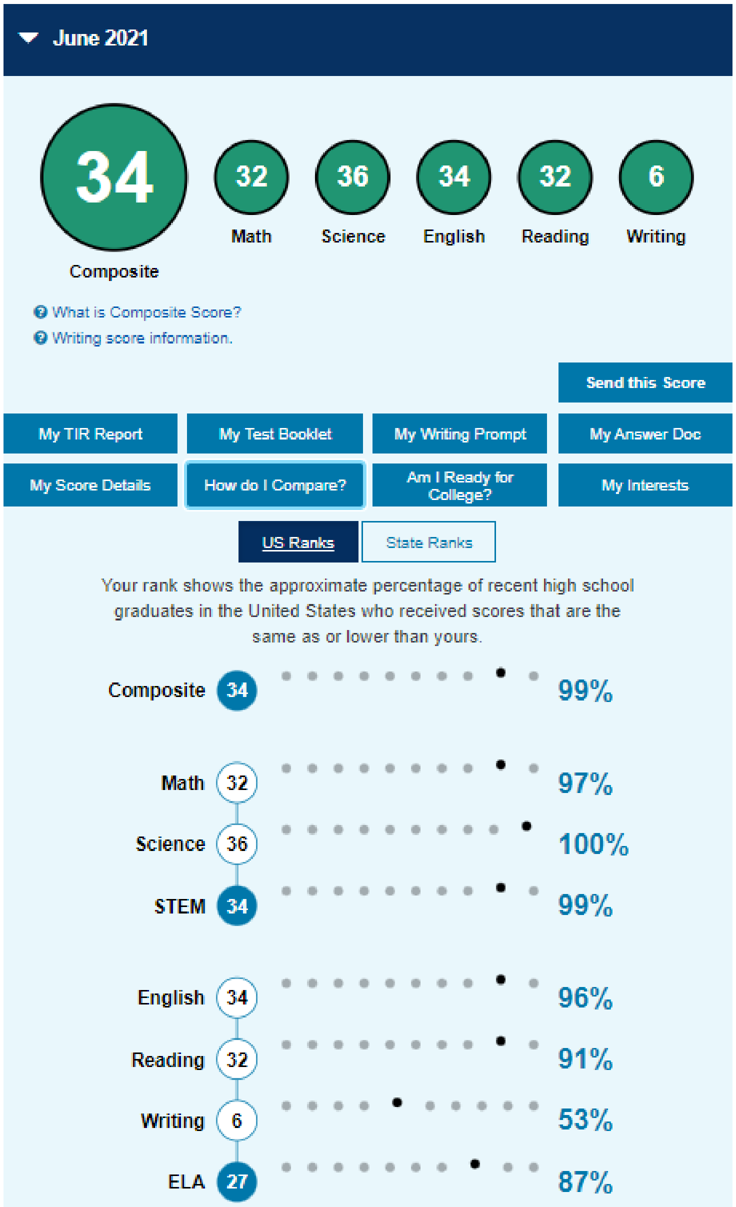
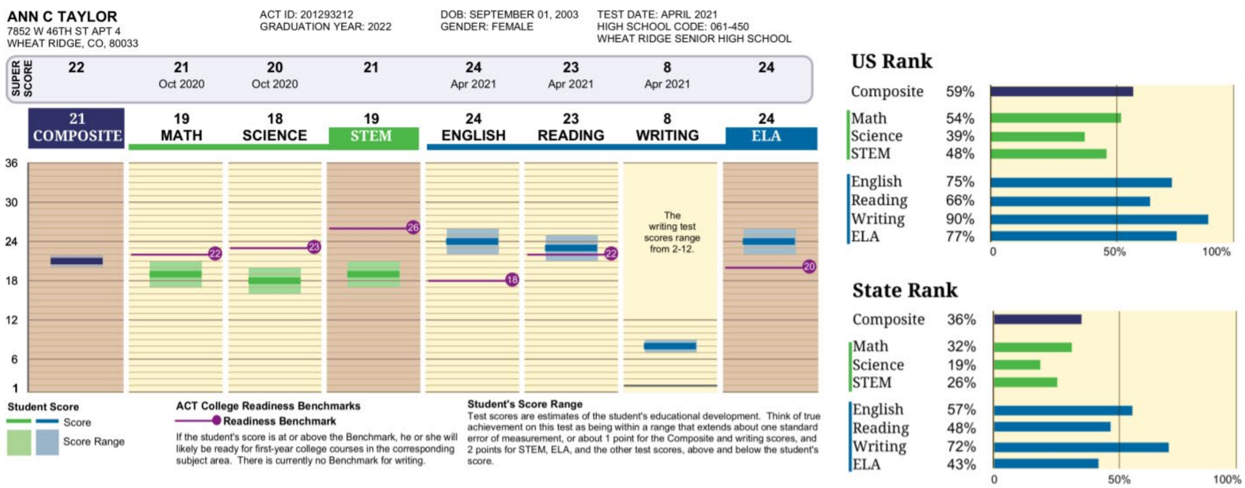


Figure 5.2. Overall Score and Percentile Rank on a Sample ACT High School Score Report



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 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 issue and its perspectives. The analysis addresses implications, complexities and tensions, and/or underlying values and assumptions.	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 enrich ideas and analysis.	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 argument. Transitions between and within paragraphs consistently clarify the relationships among ideas.	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, they do not impede understanding.
Score 4:	The writer generates an argument that	Development of ideas and support	The response exhibits a clear	The use of language conveys

Score point	Ideas & Analysis	Development & Support	Organization	Language Use & Conventions
Responses at this score point demonstrate adequate skill in writing an argumentative essay.	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	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.	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 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 argument establishes a limited or tangential context for analysis of the issue and its perspectives. Analysis is simplistic or somewhat unclear.	Development of ideas and support for claims are mostly relevant but are overly general or simplistic. Reasoning and illustration largely clarify the argument but may be somewhat repetitious or imprecise.	The response exhibits a basic organizational structure. The response largely coheres, with most ideas logically grouped. Transitions between and within paragraphs sometimes clarify the relationships among ideas.	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. 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	The writer generates an argument that weakly responds to multiple perspectives	Development of ideas and support for claims are weak, confused, or	The response exhibits a rudimentary organizational	The use of language is inconsistent and often unclear. Word

Score point	Ideas & Analysis	Development & Support	Organization	Language Use & Conventions
demonstrate weak or inconsistent skill in writing an argumentative essay	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.	disjointed. Reasoning and illustration are inadequate, illogical, or circular, and fail to fully clarify the argument.	structure. Grouping of ideas is inconsistent and often unclear. Transitions between and within paragraphs are misleading or poorly formed.	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 point demonstrate little or no skill in writing an argumentative essay.	The writer fails to generate an argument that responds intelligibly to the task. The writer's intentions are difficult to discern. Attempts at analysis are unclear or irrelevant.	Ideas lack development, and claims lack support. Reasoning and illustration are unclear, incoherent, or largely absent.	The response does not exhibit an organizational structure. There is little grouping of ideas. When present, transitional devices fail to connect ideas.	The use of language fails to demonstrate skill 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 at https://www.act.org/content/act/en/research/pdfs/R2307-CRASE-for-ACT-Writing-Technical-Report-06-2023.html.

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. 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 5.5 and 5.6 in this chapter for more details about the ACT College Readiness Benchmarks and ACT's College and Career Readiness Standards.

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 seven of the test forms administered from June 2023 to April 2024 were analyzed to obtain descriptive statistics reported in this chapter. The data set included large national samples. This part presents the summary statistics and correlations among the test section scores and the Composite, STEM, and ELA scores. Effective weights are also reported for each component of the Composite, STEM, and ELA scores. Select results are also provided for ACT International tests and ACT Superscores.

Score Distribution Summary Statistics

The summary statistics of the ACT test scores on seven of the forms administered from June 2023 to April 2024 are presented in Table 5.3. This table also includes summary statistics for multiple forms administered to ACT International test takers during the same time. Table 5.4 provides corresponding statistics for the best test section score and ACT Superscore distributions calculated with data from the same students. For students who tested only once, their single test attempts were used for superscoring. Since the ACT Superscore reflects the best test section scores for each student, the ACT Superscore means in Table 5.4 are slightly higher than those in Table 5.3.

Table 5.3. Summary Statistics of the ACT Test Score Distributions

Program	Statistic	English	Math	Reading	Science	Writing	Comp.	STEM	ELA
ACT Natl.	Mean	22.33	21.67	23.54	22.29	6.79	22.58	22.23	21.31
	SD	6.8	5.76	6.88	5.6	1.52	5.69	5.41	5.59
	Skewness	0.17	0.37	0.07	0.13	−0.17	0.14	0.27	0.02
	Kurtosis	2.31	2.24	2.12	2.75	3.05	2.24	2.40	2.38
ACT Intl.	Mean	22.99	26.41	21.34	24.39	7.26	23.9	25.64	21.75
	SD	7.74	7.12	7.48	7.29	1.77	6.76	6.88	6.07
	Skewness	0.11	−0.19	0.5	0.1	−0.09	0.14	−0.04	−0.03
	Kurtosis	1.93	1.73	2.11	1.99	3.26	1.9	1.8	2.13

Table 5.4. Summary Statistics of the Best ACT Test Section Score and ACT Superscore Distributions

Program	Statistic	English	Math	Reading	Science	Writing	Comp.	STEM	ELA
ACT Natl.	Mean	22.75	22.16	24.48	23.04	6.90	23.23	22.85	22.38
	SD	6.88	5.83	6.85	5.64	1.53	5.82	5.51	5.48
	Skewness	0.17	0.35	0.00	0.16	−0.14	0.14	0.28	−0.07
	Kurtosis	2.25	2.22	2.06	2.68	3.06	2.20	2.36	2.38
ACT Intl.	Mean	23.69	26.69	22.55	25.19	7.38	24.66	26.18	22.69
	SD	8.04	7.2	7.75	7.41	1.76	7.05	7.02	6.15
	Skewness	0.04	−0.2	0.33	0.04	−0.13	0.08	−0.07	−0.16
	Kurtosis	1.82	1.71	1.85	1.88	3.24	1.79	1.75	2.13

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 from June 2023 to April 2024 (see Table 5.5). The effective weight for the English test was computed by adding the four numbers in the first row (42.33, 26.33, 31.80, and 27.33). 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 0.29 (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, STEM, and ELA scores based on seven of the test forms administered from June 2023 to April 2024. 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 One ACT Test Form

Test	English	Math	Reading	Science
English	43.03	27.53	35.12	28.11
Math	27.53	30.19	24.91	24.29
Reading	35.12	24.91	46.64	27.52
Science	28.11	24.29	27.52	29.67

Table 5.6. Range of Effective Weights of the ACT Tests

Test	Composite	STEM	ELA
English	0.27–0.28	—	0.36–0.38
Math	0.22–0.23	0.50–0.52	—
Reading	0.27–0.28	—	0.36–0.38
Science	0.21–0.24	0.48–0.50	—
Writing	—	—	0.24–0.27

Correlations

Table 5.7 shows the correlations among the ACT test scores based on operational data from seven test forms administered from June 2023 to April 2024. 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 values of the ACT test sections are in Chapter 6.

Table 5.7. Correlations Among the ACT Test Scores

Test	English	Math	Reading	Science	Composite	STEM	Writing	ELA
English	1.00	0.77	0.80	0.78	0.93	0.81	0.53	0.91
Math	—	1.00	0.68	0.81	0.89	0.95	0.47	0.73
Reading	—	—	1.00	0.76	0.90	0.75	0.50	0.90
Science	—	—	—	1.00	0.92	0.95	0.49	0.79
Composite	—	—	—	—	1.00	0.95	0.55	0.93
STEM	—	—	—	—	—	1.00	0.50	0.80
Writing	—	—	—	—	—	—	1.00	0.75
ELA	—	—	—	—	—	—	—	1.00

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

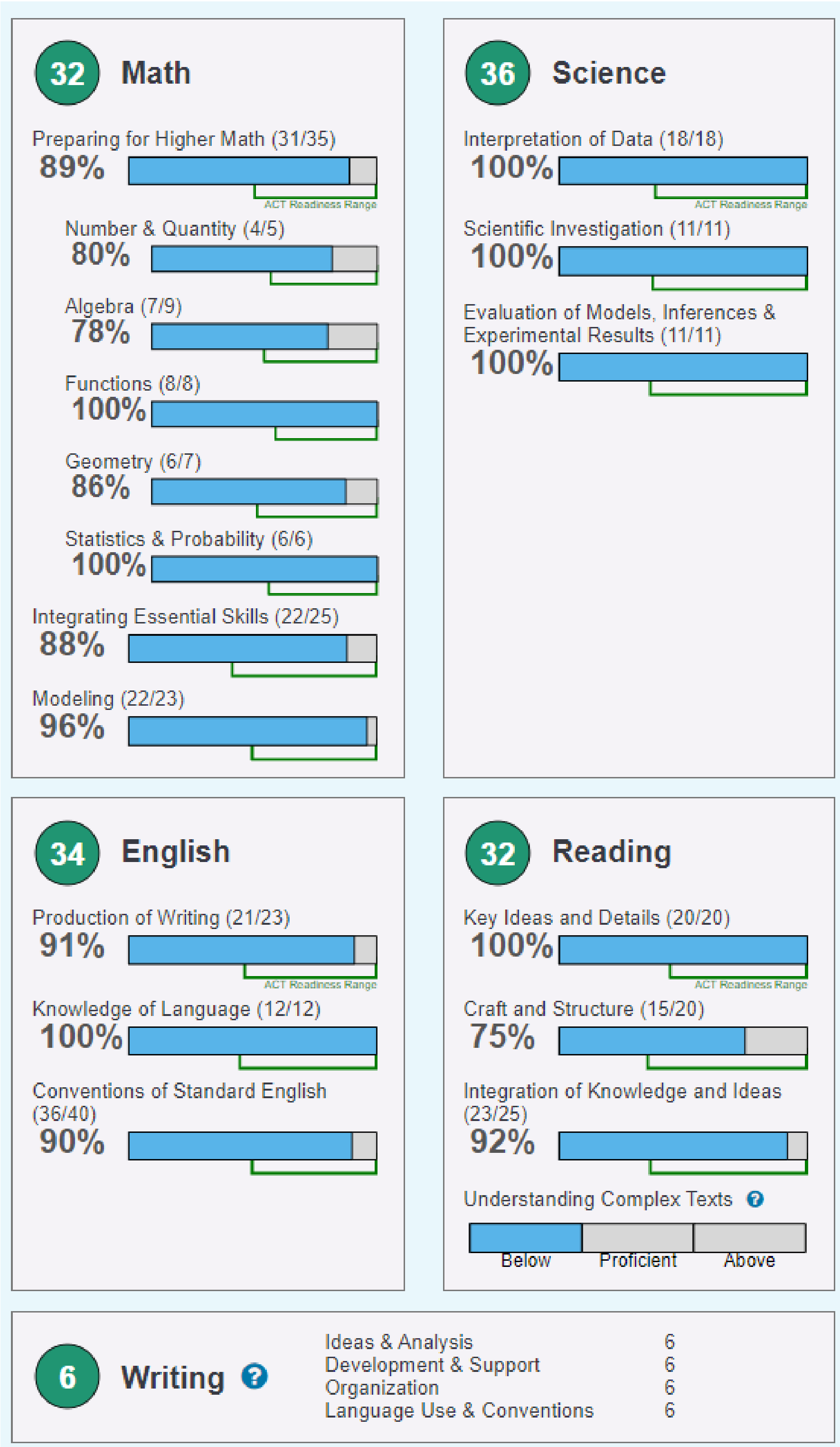
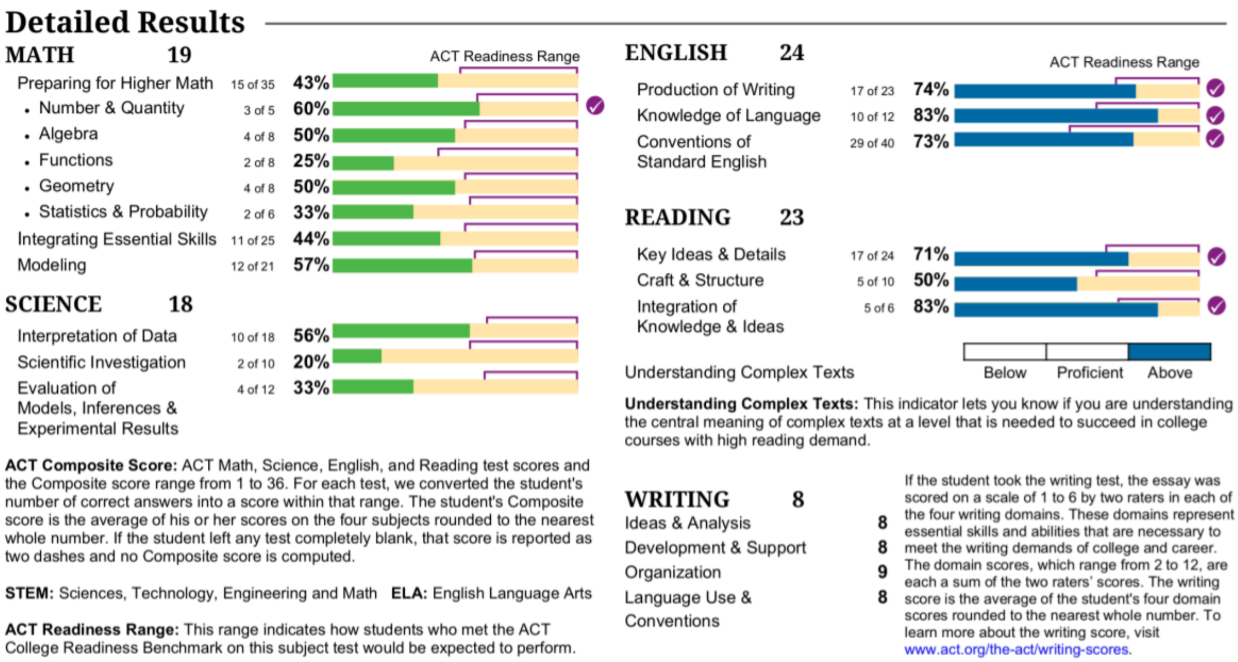


Figure 5.4. Detailed Results on a Sample ACT High School Score Report



5.3.1 Reporting Categories and ACT Readiness Ranges

ACT reporting categories are aligned with the ACT College and Career Readiness Standards (see 5.5 in this chapter) 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 section. 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 section. 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 section. 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 et al., 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 National and ACT International writing test forms administered from June 2023 to April 2024. Table 5.9 presents the correlations among these scores for ACT National testers.

Table 5.8. Summary Statistics of the ACT Writing and Writing Domain Score Distributions

Program	Statistic	Ideas & Analysis	Development & Support	Organization	Language Use & Conventions	Writing score
ACT Natl.	Mean	6.73	6.11	6.63	7.10	6.79
	SD	1.64	1.65	1.60	1.41	1.52
	Skewness	−0.27	0.03	−0.32	−0.12	−0.17
	Kurtosis	3.08	2.67	3.06	3.36	3.05
ACT Intl.	Mean	7.26	6.83	7.11	7.48	7.26
	SD	1.84	1.8	1.76	1.74	1.77
	Skewness	−0.08	0.05	−0.13	−0.1	−0.09
	Kurtosis	3.22	3.05	3.24	3.44	3.26

Table 5.9. Correlations Among the ACT Writing and Writing Domain Scores

Score	Ideas & Analysis	Development & Support	Organization	Language Use & Conventions	Writing score
Ideas & Analysis	1.00	0.90	0.97	0.92	0.97
Development & Support	—	1.00	0.90	0.86	0.91
Organization	—	—	1.00	0.90	0.97
Language Use & Conventions	—	—	—	1.00	0.95
Writing score	—	—	—	—	1.00

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 et al., 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[1]: 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; and

[1] 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 5.6 in this chapter).

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

Item no.	Score range					
	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
Math	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 8th, 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

Test	Task 1 (%)	Task 2 (%)
English	75	86
Math	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 (Table 5.1 in 5.2.2 in this chapter). 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, 2016c). 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	Math	22
American history, other history, psychology, sociology, political science, & economics	Reading	22
Biology	Science	23
Calculus I, biology, chemistry, physics, & engineering	STEM	26
English Composition I & 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 et al., 2015), and the third study developed the ELA Benchmark (Radunzel et al., 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 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 (1989).

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 (Kolen, 1991; Kolen & Hanson, 1989; Lord, 1965), 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). 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.

6.1.5 IRT Equating for ACT International Testing

As of September 2018, all international administrations of the ACT are delivered via laptops and desktops using PSI's ATLAS Cloud® test delivery platform. Online testing affords the opportunity to report scores quickly because it is unnecessary to ship, scan, and score answer documents. To facilitate rapid score reporting, raw-to-scale score conversion tables for the English, math, reading, and science tests are generated in advance of testing through the process of item response theory (IRT) true-score equating. Writing scores cannot be reported quickly because each response must be scored by at least two raters.

IRT equating begins with “calibrating” the items with BILOG-MG 3.0 (Zimowski et al., 2003) using data from ACT equating events. This involves estimating three parameters for each item: the difficulty parameter (b_j), the discrimination parameter (a_j), and the pseudo-chance (or “guessing”) parameter (c_j) indicating the probability that a low ability examinee will choose the correct answer. These three item parameters define the relationship between examinee ability θ_i and the probability of responding correctly to an item in the 3-parameter logistic (3PL) IRT model (Birnbaum, 1968). The 3PL model is defined as

$$P\left(x_{ij} = 1 \middle| \theta_i, a_j, b_j, c_j\right) = c_j + \frac{1 - c_j}{1 + e^{-Da_j(\theta_i - b_j)}}$$

where x_{ij} is the item score (0 or 1) of examinee i to item j, D is a scaling constant equal to 1.702, and θ_i is the examinee latent trait (achievement in the case of the ACT). When plotted, this S-shaped curve is known as the item characteristic curve (ICC).

The IRT parameters are then transformed to the same scale as all previously calibrated items using the Stocking-Lord method (Stocking & Lord, 1983). Note that nearly all items in the item bank have IRT parameters based on data from a paper administration. Considering mode effects between paper and online testing (Section 6.4), those parameters would not be appropriate for use in generating a raw-to-scale score conversion table for an ACT form administered online (like all ACT International forms). An IRT-based mode adjustment was estimated using data from the 2018 mode comparability study (Section 6.4.4). Specifically, the Stocking-Lord method was used to estimate the relationship between item parameters for paper and online versions of the same items. This same transformation can be applied to any item to estimate parameters appropriate for an online test administration.

With all item parameters on the same scale (appropriate for online administration), it becomes possible to equate test forms made up of any combinations of items from the IRT-calibrated item bank. When a form is developed for online international testing, it is equated to a base form with a pre-existing raw-to-scale score conversion table (the form administered online in the 2018 mode comparability study; see Section 6.6.4). The equating process involves generating the test characteristic curves (TCCs) for the new form and the base form. A TCC, which is simply the sum of the ICCs, shows the relationship between examinee ability and expected raw score (number correct) on a form. IRT true-score equating treats as equivalent the raw scores on two different forms that correspond to the same value on the latent trait (θ) scale. Finally, the equated raw scores and the raw-to-scale score conversion table of the base form are used to obtain the raw-to-scale score conversion table of the new form.

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:

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

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:

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

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 $person \times (items{:}\,content)$ 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 test sections (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 seven of the test forms administered from June 2023 to April 2024. The reliability estimates were high, with values of 0.92 or greater for English, mathematics, Composite, STEM, and ELA scores, and values of 0.83 or greater for reading and science. By design, the SEM should be about 1 for the Composite score and about 2 for the section tests. Reliability and SEM values were fairly consistent across forms.

Table 6.1. Summary Statistics of Scale Score Reliability and SEM for the ACT Test Scores

Test	No. of items	Reliability			SEM		
		Mdn	Min	Max	Mdn	Min	Max
English	75	0.94	0.92	0.94	1.68	1.64	1.82
Math	60	0.92	0.92	0.93	1.53	1.51	1.65
Reading	40	0.88	0.85	0.89	2.36	2.28	2.62
Science	40	0.88	0.83	0.89	2.01	1.90	2.13
Composite	215	0.97	0.96	0.97	0.97	0.94	1.02
STEM	100	0.95	0.93	0.95	1.26	1.22	1.31
ELA	116	0.93	0.92	0.94	1.44	1.43	1.50

Note. Mdn = median; Min = minimum; Max = maximum

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 seven forms administered from June 2023 to April 2024. 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. Summary Statistics of Raw Score Reliability and SEM for the ACT Reporting Categories

Test	Reporting Categories	No. of items	Reliability			SEM		
			Mdn	Min	Max	Mdn	Min	Max
English	Production of Writing	23	0.84	0.82	0.87	1.92	1.84	2.07
	Knowledge of Language	12	0.72	0.68	0.79	1.32	1.10	1.44
	Conventions of Standard English	40	0.90	0.88	0.91	2.80	2.40	2.87
Math	Preparing for Higher Mathematics	36	0.89	0.86	0.90	2.52	2.45	2.52
	Number & Quantity	6	0.55	0.50	0.61	1.08	0.95	1.08
	Algebra	8	0.68	0.53	0.72	1.20	1.12	1.20
	Functions	8	0.66	0.56	0.71	1.20	1.12	1.28
	Geometry	8	0.61	0.57	0.73	1.20	1.12	1.20
	Statistics & Probability	6	0.55	0.50	0.61	1.02	0.96	1.08
	Integrating Essential Skills	24	0.85	0.83	0.86	2.16	1.92	2.25
	Modeling	21	0.81	0.77	0.83	1.89	1.60	2.10
Reading	Key Ideas & Details	23	0.80	0.78	0.82	2.07	1.98	2.16
	Craft & Structure	11	0.71	0.62	0.72	1.43	1.30	1.44
	Integration of Knowledge & Ideas	6	0.55	0.40	0.59	1.08	0.76	1.19
Science	Interpretation of Data	18	0.78	0.67	0.80	1.80	1.71	1.90
	Scientific Investigation	10	0.63	0.50	0.72	1.30	1.20	1.40
	Evaluation of Models, Inferences & Experimental Results	12	0.72	0.63	0.78	1.44	1.43	1.68

Note. Mdn = median; Min = minimum; Max = maximum

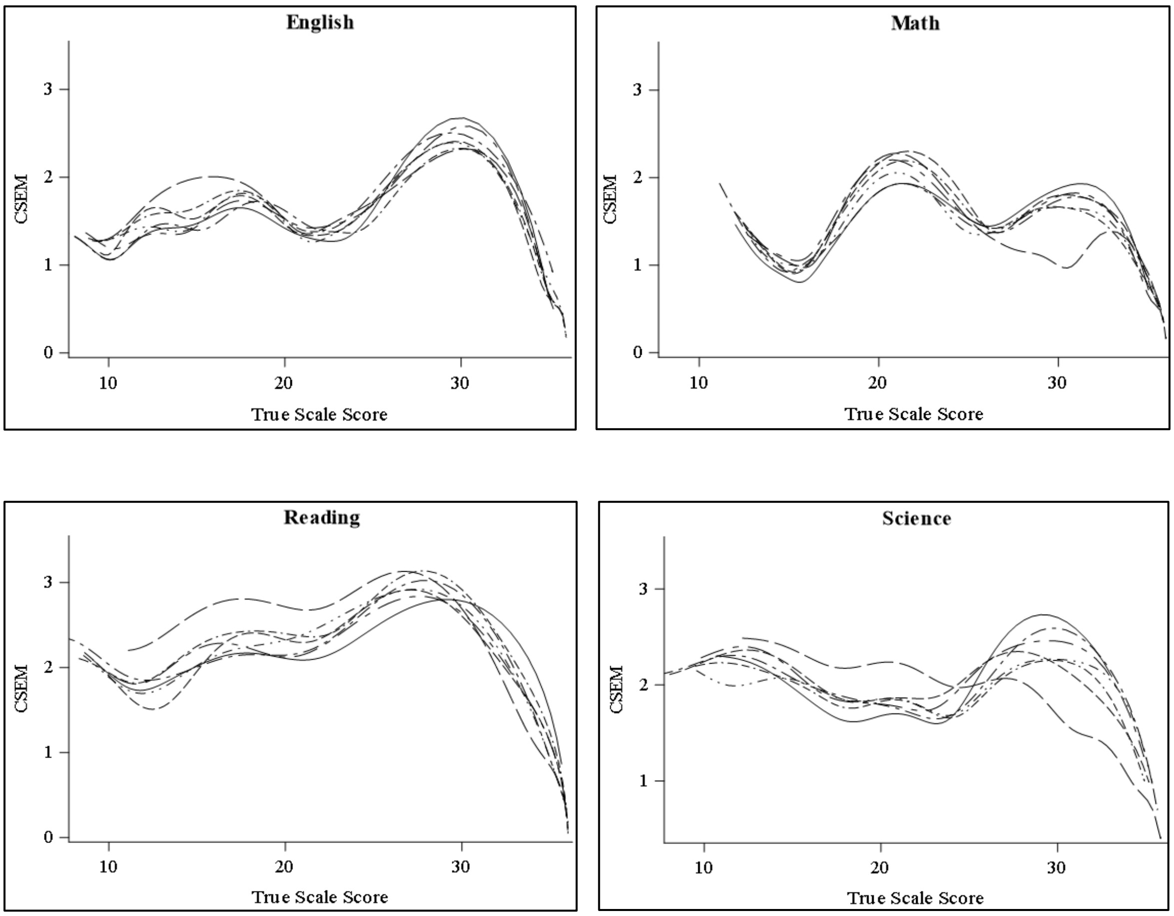
Conditional Standard Errors of Measurement for the ACT Multiple-Choice Test Scores

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 test sections for seven of the forms administered from June 2023 to April 2024. 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 test sections 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, math, and reading tests had greater variation along the score scale, but in most of the true scale score range, the CSEM is about 2. For all test sections, 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

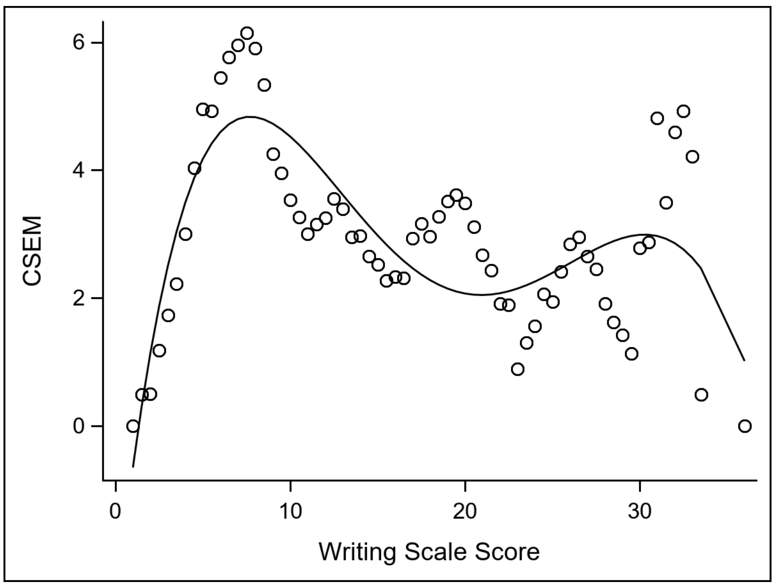


Reliability, CSEM, and Agreement Indices for the ACT Writing Test Scores

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 also calculated based on seven forms administered from June 2023 to April 2024. This included the perfect agreement rate, the perfect plus adjacent agreement rate, and the quadratic weighted kappa coefficient (Table 6.3). The perfect agreement rate, or percentage of students who received the same domain score (from 1 to 6) from both raters, ranged from 0.690 to 0.763 across domains and forms. The perfect plus adjacent agreement rates, or the percentage of students who received either the same domain score or adjacent domain scores (e.g., a score of 5 and a score of 6) from the two raters, was very high, ranging from 0.994 to 0.999 across domains and forms.

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 (2003) 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 0.695 to 0.836, indicating good rater agreement.

Table 6.3. Agreement Rates for the ACT Writing Domain Scores

Domain	Agreement index	Median	Minimum	Maximum
Ideas & Analysis	Perfect agreement	0.719	0.695	0.761
	Perfect + adjacent agreement	0.996	0.994	0.999
	Quadratic weighted kappa	0.801	0.765	0.836
Development & Support	Perfect agreement	0.723	0.691	0.754
	Perfect + adjacent agreement	0.997	0.995	0.999
	Quadratic weighted kappa	0.800	0.771	0.834
Organization	Perfect agreement	0.731	0.698	0.763
	Perfect + adjacent agreement	0.998	0.996	0.999
	Quadratic weighted kappa	0.800	0.754	0.831
Language Use & Conventions	Perfect agreement	0.724	0.690	0.759
	Perfect + adjacent agreement	0.997	0.994	0.999
	Quadratic weighted kappa	0.733	0.695	0.775

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

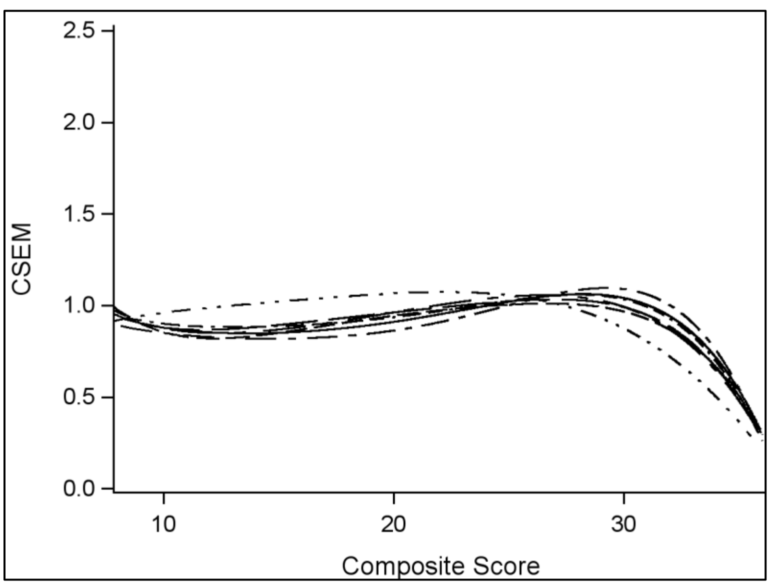
$$s_c(\tau_e, \tau_m, \tau_r, \tau_s) = \frac{\sqrt{\sum_i s_i^2(\tau_i)}}{4}, \tag{3}$$

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 test sections were treated as true proportion-correct scores at which the CSEMs were calculated. For each section, 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 section, 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 test sections 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 seven test forms administered in June 2023 to April 2024 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 section scores rather than the rounded average of the four test section 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

$$S_{STEM}(\tau_m, \tau_s) = \frac{\sqrt{\sum_i s_i^2(\tau_i)}}{2}, \tag{4}$$

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

$$S_{ELA}(\tau_e, \tau_r, \tau_w) = \frac{\sqrt{\sum_i s_i^2(\tau_i)}}{3}, \quad (5)$$

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

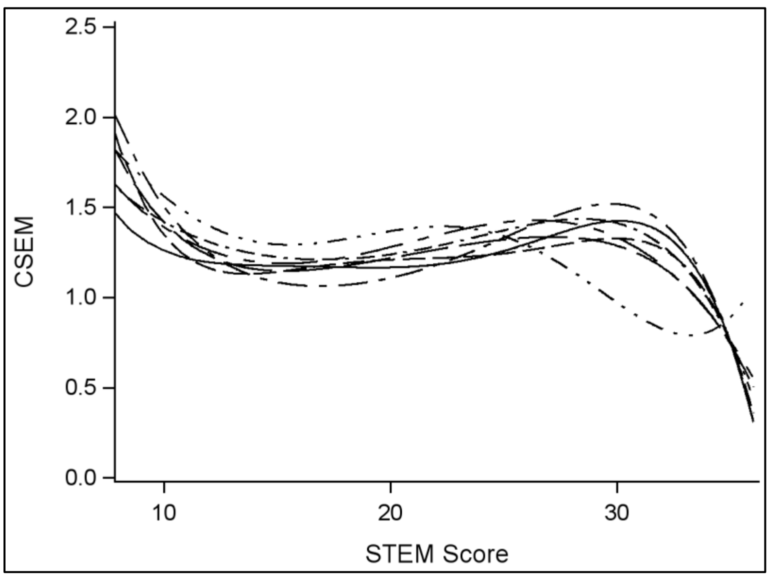
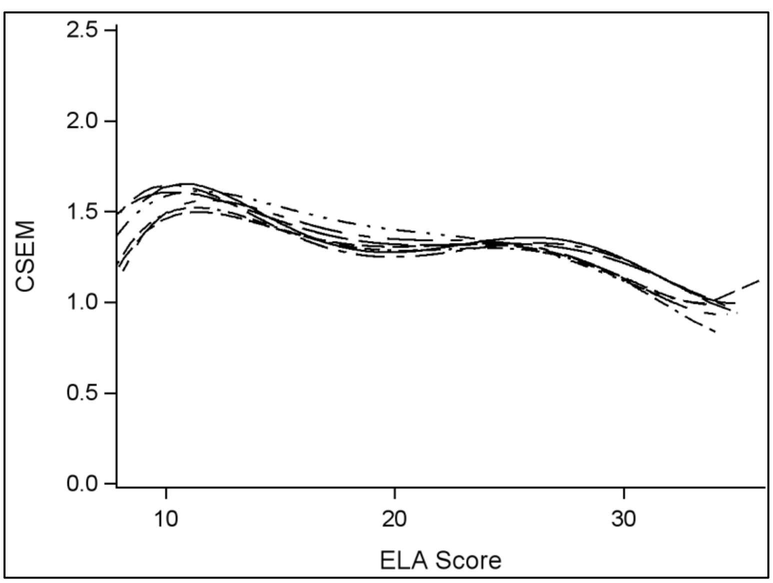


Figure 6.5. CSEM for ELA Scores



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. Below are classification consistency results for the ACT test scores and indicators.

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

Classification consistency values were computed using data from seven forms administered in from June 2023 to April 2024 for the four ACT multiple-choice tests and the STEM and ELA scores. Classification was based on the ACT College Readiness Benchmarks (see Chapter 5 for details about the Benchmarks). The classification consistency results are provided in Table 6.4. Values are all fairly high, ranging from a low of 0.83 for science to a high of 0.94 for STEM.

Table 6.4. Classification Consistency for the ACT Readiness Benchmarks

		Classification consistency		
Test	Number of items	Median	Minimum	Maximum
English	75	0.90	0.89	0.91
Math	60	0.90	0.89	0.91
Reading	40	0.86	0.85	0.87
Science	40	0.86	0.83	0.87
STEM	100	0.93	0.91	0.94
ELA	116	0.89	0.88	0.90

Similarly, classification consistency for the ACT Readiness Ranges was computed for each of the ACT test reporting categories. These values, provided in Table 6.5, are based on data from seven forms administered from June 2023 to April 2024.

Table 6.5. Classification Consistency for the ACT Readiness Ranges

Test	Reporting Categories	No. of items	Reliability		
			Median	Minimum	Maximum
English	Production of Writing	23	0.87	0.84	0.89
	Knowledge of Language	12	0.82	0.76	0.90
	Conventions of Standard English	40	0.89	0.86	0.91
Math	Preparing for Higher Mathematics	36	0.86	0.85	0.87
	Number & Quantity	6	0.70	0.68	0.73
	Algebra	8	0.75	0.68	0.80
	Functions	8	0.74	0.71	0.79
	Geometry	8	0.73	0.69	0.78
	Statistics & Probability	6	0.68	0.63	0.74
	Integrating Essential Skills	24	0.84	0.83	0.85
	Modeling	21	0.81	0.80	0.83
Reading	Key Ideas & Details	23	0.82	0.79	0.83
	Craft & Structure	11	0.77	0.72	0.78
	Integration of Knowledge & Ideas	6	0.71	0.65	0.73
Science	Interpretation of Data	18	0.80	0.75	0.81
	Scientific Investigation	10	0.73	0.68	0.77
	Evaluation of Models, Inferences & Experimental Results	12	0.77	0.73	0.81

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 June 2023 to April 2024, the classification consistency ranged from 0.64 to 0.68, 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 20 across these seven forms, and the percentages of students classified as Below Proficient, Proficient, and Above Proficient were 41%, 29%, and 30%, respectively.

Classification Consistency for Progress Toward WorkKeys NCRC Indicator

The second indicator, Progress Toward the ACT® WorkKeys® National Career Readiness Certificate® (NCRC®), had classification consistency values ranging from 0.76 to 0.82 across seven of the forms administered from June 2023 to April 2024. These values are quite high considering that there are four performance levels for the WorkKeys NCRC, as shown in Table 6.6. Note that the classification consistency index is an indication of the stability of the Progress Toward the WorkKeys 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 WorkKeys National Career Readiness Certificate Indicator.

Table 6.6. Composite Score Ranges for the WorkKeys NCRC Levels

WorkKeys NCRC level	Composite score range
Unlikely to earn a WorkKeys NCRC	1–12
Most likely to obtain a Bronze level	13–16
Most likely to obtain a Silver level	17–21
Most likely to obtain a Gold level	22–26
Most likely to obtain a Platinum level	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. In February 2024, online testing was introduced for the national testing program.

As of spring 2025, the ACT may be administered on paper or online for both national testing and state and district testing but only online for international testing. 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 uses the PSI UTD platform for online testing in its national testing program. Test centers with online capability will provide test center–managed devices. At the present time, testing is supported on Chromebooks and Windows desktops and laptops.

ACT collaborated with Pearson to design the TestNav platform architecture for the ACT online test delivery system for state and district testing. Test centers can use this test delivery system across multiple device types, including laptop and desktop computers running operating systems 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 using experimental design or quasi-experimental design. For national testing and state and district testing, subsequent online forms are equated to the online base forms through online test equating studies based on random groups or through separate mode studies based on matched samples. 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.

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 evaluate and support use of the TAO platform for online ACT delivery. Finally, a mode study was conducted in December 2023 to support the launch of online testing for the national testing program using the PSI UTD platform.

All studies except the December 2023 mode study 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 the 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 section 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.

2018 ATLAS Cloud Study

To enhance test security and to provide faster score reporting, paper administrations of the ACT have been discontinued in international administrations. Thus, as of September 2018, all international administrations of the ACT are delivered via laptops and desktops using PSI's ATLAS Cloud test delivery software. Prior studies examined comparability between paper and online ACT testing on the TestNav platform (e.g., Li, Yi, & Harris, 2017), but items are displayed differently in ATLAS Cloud, and this could lead to different mode effects. For that reason, ACT conducted a mode comparability study in 2018 with participants randomly assigned to one of two testing conditions: online or paper testing on ATLAS Cloud. Since the groups testing on paper and online were randomly equivalent, observed differences in performance were attributed to mode differences, and statistical adjustments were used to eliminate mode differences such that scores from either mode represented the same level of performance.

Analyses in the 2018 ATLAS Cloud mirrored those in the preceding TestNav studies. Results revealed very small differences between paper and online testing in terms of correlations among tests, effective weights, and reliability for the English, mathematics, reading, and science tests. However, average item scores and test scores were slightly higher on average for students who tested online compared to those who tested on paper, especially on the English and reading tests. Mathematics and science scores were relatively more comparable between the two administration modes. The average differences, in standard deviation units, were 0.16 for English, 0.05 for math, 0.24 for reading, and 0.07 for science. The English and reading differences were statistically significant ($p < .001$), and the science difference was nearly so ($p = .06$). In general, results indicated that the mode effects on ATLAS Cloud were similar in magnitude to those observed on TestNav. On the ACT writing test, scores tended to be higher for students who took the online version compared those who took the paper version. Equating methodology was applied to adjust for the differences so that scale scores from online and paper testing were comparable. Once an adjustment was made to the study forms, it was carried forward to future operational forms using item response theory (IRT) equating methods (see Section 6.1.5).

2019–2020 TAO Mode Comparability Studies

Students who register for a Saturday national testing event have the option to test online. In 2019, students were expected to take the ACT on the TAO platform instead of the Pearson TestNav platform used for state and district testing. Due to concerns that test scores from different online testing platforms might exhibit different mode effects, a series of mode comparability studies was conducted during the 2019–2020 academic year for online test delivery through TAO (Steedle, Pashley, and Cho, 2020). Although TAO is no longer used in national testing, the results from the three mode studies are briefly described below.

Three studies took place on the Saturday national testing dates in October 2019, December 2019, and February 2019. Only the February 2019 study included writing as an optional component. As in earlier mode comparability studies, students were randomly assigned to test on paper or online, and all participants received college reportable scores. In each study, the same form was administered on paper and online, but a different form was used for each study. The analyses of construct equivalency and score equivalency were the same as those used in the spring 2014 and spring 2015 studies.

In general, the results were quite consistent across studies and with prior ACT mode comparability studies. The construct equivalency analyses indicated that paper and online testing appeared to be comparable in terms of correlations among the subject areas, effective weights, internal consistency reliability, and confirmatory factor analysis model fit and average factor loadings. The score equivalence analyses revealed that online scores were higher than paper scores on average, especially on the reading and English tests. Across studies, the mode effect ranged from 0.16 to 0.22 in reading and from 0.10 to 0.13 in English. The mode effects ranged from 0.04 (nonsignificant) to 0.12 in science, and they ranged from −0.01 (nonsignificant) to 0.06 in math. In all cases, the online test was equated to the paper test to ensure that scores reported from this study would be comparable regardless of testing mode.

2023 UTD Mode Comparability Study

Following a decision to adopt the UTD testing platform for the national testing program, an additional mode study was conducted on the Saturday national test date in December 2023. This study included both the multiple-choice tests and writing tests and utilized propensity score matching to investigate mode effects.

Results from the UTD mode study were also similar to findings from earlier mode studies. Little differences were found between paper and online in terms of correlations among the subject areas, effective weights, internal consistency reliability, and confirmatory factor analysis model fit and average factor loadings. Yet at least some test sections exhibited significant mode differences in score distributions, especially the reading and writing tests. The effect sizes of online versus paper mode differences for English, math, reading, science, and writing were 0.01 (nonsignificant), −0.03 (nonsignificant), 0.23, 0.10, and 0.37, respectively. The online test was equated to the paper test to ensure that scores reported from the mode studies would be comparable regardless of testing mode.

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 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

---

[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).

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 mathematics 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.

7.1.3 Statistical Relationships between High School Coursework and Grades and ACT Scores

The ACT tests are oriented toward the general content areas of high school and college curricula. In particular, ACT conducts a National Curriculum Survey every three to five years to ensure that our assessments always measure the knowledge and skills being taught in schools around the country (ACT, 2020b; see Chapter 2). Students' performance on the ACT should therefore be related to the high school courses they have taken and to their performance in these courses.

One component of registering for the ACT is completion of the Course/Grade Information Section (CGIS), which collects information on about 30 high school courses in English, mathematics, social studies, natural sciences, languages, and arts. Many of these courses form the basis of a high school college-preparatory curriculum and are frequently required for college admission or placement. For each of the 30 courses, students indicate whether they have taken or are currently taking the course, whether they plan to take it, or do not plan to take it. If they have taken the course, they indicate the grade they received (A–F). Self-reported coursework and grades collected with the CGIS have been found to be accurate relative to information provided on student transcripts (Sanchez & Buddin, 2016; Sawyer, Laing, & Houston, 1988; Valiga, 1986; see also the next section).

Table 7.1 displays the ACT scale score means and the percentage of students meeting the College Readiness Benchmarks for the English, mathematics, reading, and science tests by common high school course-taking patterns in English, mathematics, social studies, and science. As seen in the table, mean test scores and the percentage of students meeting the Benchmarks typically increase as the years of coursework increases in that subject area. However, the strength (and sometimes direction) of this relationship is also conditional on the types of courses taken. For example, students who take mathematics courses through Calculus and science courses through Physics typically have higher mean scores and Benchmark attainment rates than students who complete a different sequence of courses over the same number of years of coursework.

Table 7.1. College Readiness Benchmark Attainment Percentages and Average ACT Scores by Common Course Patterns, ACT-Tested High School Graduating Class of 2021

Course pattern		*N*	Mean ACT score	Percentage (%) meeting benchmark
English	Eng 9/Eng 10/Eng 11/Eng 12/other	165,958	22.3	72
	Eng 9/Eng 10/Eng 11/Eng 12	623,323	21.1	65
	Less than 4 years of English	56,980	16.6	38
	Zero years/no English	449,088	16.8	39
Mathematics	Alg 1/Alg 2/geom/trig/calc	41,584	23.2	61
	Alg 1/Alg 2/geom/trig/other adv math	58,940	21.4	49
	Alg 1/Alg 2/geom/trig	38,353	18.8	26
	Alg 1/Alg 2/geom/other adv math	179,211	19.1	28
	Other comb of 4 or more years of math	326,953	24.1	66
	Alg 1/Alg 2/geom	98,409	16.5	9
	Other comb of 3 or 3.5 years of math	50,170	19.5	32
	Less than 3 years of math	44,627	16.0	9
	Zero years/no math	457,102	18.1	22
Social Studies	U.S. Hist/World Hist/Am Gov/other hist	25,929	22.6	53
	Other comb of 4 or more years soc sci	400,925	23.3	57
	U.S. Hist/ World Hist/Am Gov	63,786	19.9	37
	Other comb of 3 or 3.5 years of soc sci	247,455	22.1	50
	Less than 3 years of soc sci	99,996	19.8	36
	Zero years/no soc sci	457,258	18.5	30
Science	Gen sci/bio/chem/phys	326,363	22.9	52
	Bio/chem/phys	109,874	24.0	60
	Gen sci/bio/chem	244,298	20.4	33
	Other comb of 3 years of nat sci	25,088	19.2	26
	Less than 3 years of nat sci	131,559	18.5	22
	Zero years/no nat sci	458,167	18.5	23

Moreover, as shown in Table 7.2, students who have completed or plan to complete a core curriculum tend to achieve higher ACT scores than those who have not completed a core curriculum (ACT, 2016b), where a core curriculum is defined by at least four years of English and at least three years each of mathematics, social studies, and natural sciences. From 2011–2012 through 2015–2016, the ACT Composite scores of students who completed a core curriculum averaged about 3 scale score points higher than the scores of those who did not.

Table 7.2. Average ACT Scores by Academic Preparation, 2012–2016

Academic preparation	Year	*N*	ACT score				
			English	Math	Reading	Science	Composite
Core curriculum or more completed	2011–12	1,259,744	21.3	21.8	22.0	21.6	21.8
	2012–13	1,322,739	21.2	21.7	22.0	21.5	21.7
	2013–14	1,347,997	21.4	*21.7*	*22.2*	21.6	21.8
	2014–15	1,389,338	21.4	21.7	22.3	21.8	21.9
	2015–16	1,441,538	21.3	21.5	22.3	21.7	21.9
Core curriculum not completed	2011–12	355,849	18.3	19.1	19.4	19.1	19.1
	2012–13	396,592	17.8	18.9	19.0	18.8	18.7
	2013–14	405,073	17.9	18.9	19.2	18.9	18.9
	2014–15	424,562	18.0	18.9	19.3	19.0	18.9
	2015–16	483,335	17.8	18.7	19.2	18.8	18.7

Note. Core curriculum is defined as four or more years of high school English and three or more years each of high school mathematics, social studies, and natural sciences.

The findings shown in Tables 7.1 and 7.2 are consistent with the notion that the ACT is a curriculum-based test. Additionally, an analysis by McNeish, Radunzel, and Sanchez (2015) showed that, in general, coursework and high school grades were strongly associated with performance on the ACT, after statistically controlling for other factors. However, it is also conceivable that other factors, such as noncognitive characteristics of students, account for the observed association between high school coursework and ACT scores. In the McNeish et al. study, the researchers investigated the relationships between noncognitive characteristics, high school coursework and grades, school characteristics, and ACT test scores. They found that between 44% and 61% of the variance in ACT scores was explained by high school grade point average (HSGPA), coursework taken, school characteristics, noncognitive characteristics, and demographic characteristics. High school academic factors, such as HSGPA, courses taken, and advanced coursework, accounted for the most variance explained in all five ACT scores (R^2 = 0.28 to 0.46). These three factors accounted for 64% to 77% of the total variance explained by the models. In particular, taking higher-level mathematics and science courses and subject-relevant accelerated, advanced, honors, or dual-enrollment courses were associated with sizable mean ACT score differences. Specific English and social studies courses were not included in the models because of the limited variability in students' course taking in these subject areas and their collinearity with other variables, such as coursework taken in mathematics and science. The findings from this study are consistent with earlier studies (Noble, Davenport, Schiel, & Pommerich, 1999a, b; Noble & McNabb, 1989; Schiel, Pommerich, & Noble, 1996) that examined relationships among coursework, grades, and ACT scores.

ACT research has shown that taking rigorous, college-preparatory mathematics courses is associated with higher ACT mathematics and Composite scores (e.g., ACT, 2016c; Noble, Davenport, & Sawyer, 2001; Noble, Roberts, & Sawyer, 2006). Schiel et al. (1996) statistically controlled for prior achievement using ACT Plan® scores and found substantive increases in average ACT mathematics and science scores associated with taking higher-level mathematics and science courses. In other studies, researchers found that, in a typical high school, students who take higher-level mathematics or science courses (e.g., Trigonometry, Calculus, Chemistry, or Physics) can expect to earn meaningfully higher average ACT mathematics and science scores than students who do not take such courses (Noble & Schnelker, 2007; ACT, 2005). The expected benefits of coursework taken in high school for increasing ACT performance depend on the high school students attend, regardless of prior achievement and grade level at testing (Noble & Schnelker, 2007).

7.1.4 Statistical Relationships between High School Coursework and Grades and ACT Benchmark Attainment

To provide students and educators with an empirical definition of what it means to be academically ready for first-year credit-bearing college courses, ACT developed the ACT College Readiness Benchmarks based on college course grade data from 214 two- and four-year institutions (Allen, 2013). The ACT College Readiness Benchmarks are scores on the ACT multiple-choice tests that represent the level of achievement required for students to have at least a 50% chance of obtaining a B or higher grade in related first-year college courses. The Benchmarks also correspond to an approximate 75% chance of earning a C or higher grade in these courses. The Benchmarks corresponding to the four ACT multiple-choice test scores linked to common first-year courses include: ACT English to English Composition I, ACT mathematics to College Algebra, ACT reading to social science courses, and ACT science to Biology. The Benchmarks correspond to scores of 18, 22, 22, and 23 on the ACT English, mathematics, reading, and science tests, respectively. For more details on the development of the ACT College Readiness Benchmarks, as well as details on the ACT STEM and ELA Benchmarks, see Chapter 5.

A study by Ling and Radunzel (2017) examined how the high school coursework taken and grades earned related to students' chances of meeting the ACT College Readiness Benchmarks in each of the four sections, after accounting for other student and school characteristics. The study findings indicated that students who took rigorous courses in high school and earned good grades were more likely to meet the Benchmarks and therefore were more likely to experience success in first-year college courses. These study findings are consistent with those from an earlier ACT study by Noble and Schnelker (2007). Findings from the 2007 study indicated that some courses and course sequences were more strongly associated with Benchmark attainment than others. Each incremental college-preparatory course taken, particularly in mathematics and science (e.g., Trigonometry beyond Algebra 2, Physics beyond Chemistry), increased the likelihood of meeting the Benchmarks more than the number of courses taken in a discipline alone. A limitation of these studies is that students' self-reported courses taken and grades earned are based only on those courses available in the ACT CGIS, which does not provide more detailed information on the courses taken, especially in English.

7.1.5 Distinction between ACT Scores and HSGPA as Measures of Educational Achievement

ACT scores are statistically associated with high school grades (Table 7.3; see also the previous section). That is, students who have higher HSGPAs tend to achieve higher ACT scores. As shown in Table 7.3, students in the highest HSGPA category (3.50–4.00) tend to earn ACT Composite scores nearly 10 scale score points higher than students in the lowest HSGPA category (0.00–1.99). However, ACT scores and HSGPAs are different measures in that there are some noncognitive predictors related to high school grades that are not directly related to ACT scores (McNeish et al., 2015; Noble et al., 1999a, 1999b). To the extent that grades measure educational achievement, there will be a strong statistical relationship between grades and ACT scores. However, grades are more subjective than standardized test scores because of differing grading standards and practices (Pilcher, 1994; Brookhart, 1993; Stiggins, Frisbie, & Griswold, 1989). Within a given school, teachers may differ in the criteria they use to judge student achievement. Effort and reward are often confounded with academic accomplishment in assigning course grades (Allen, 2005; Pilcher, 1994; Willingham, Pollack, & Lewis, 2002). In a review of the literature on elementary and high school grading practices over the past century, Brookhart (2015) concluded that “Report card grades can be reliable and valid measures of academic achievement, but may not be depending on individual teachers’ grading practices” (p. 268). Grading practices also vary across schools; an “A” in one school may be equivalent to a “C” in another school (United States Department of Education, 1994). Consequently, the interpretation of high school grades should take into account differences across high schools in their curricula and grading standards. Grade inflation also adversely affects the usefulness of high school grades in terms of understanding educational achievement.

Table 7.3. Average ACT Score by HSGPA Ranges, 2015–2016

HSGPA group	*N*	English		Math		Reading		Science		Composite	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
3.50–4.00	726,643	24.5	6.0	24.2	5.2	25.1	5.9	24.2	5.1	24.6	5.0
3.00–3.49	479,292	19.5	5.5	19.8	4.4	20.7	5.5	20.3	4.5	20.2	4.4
2.50–2.99	274,467	16.9	5.0	17.7	3.6	18.4	5.1	18.3	4.3	18.0	3.9
2.00–2.49	154,002	15.1	4.6	16.5	3.0	16.8	4.7	16.8	4.1	16.4	3.5
0.00–1.99	75,255	13.6	4.3	15.7	2.5	15.4	4.2	15.5	3.9	15.2	3.1
All	2,090,342	20.1	6.8	20.6	5.4	21.3	6.5	20.8	5.6	20.8	5.6

Accuracy of Self-Reported Coursework and Grades

The accuracy of the high school course and grade information students provide in the ACT registration folder within the CGIS is a focus of continuing research at ACT. Sanchez and Buddin (2016) concluded that students’ self-reported grade information accurately represented students’ high school experience. About 94% of students accurately reported taking particular courses. The correlation between self-reported and transcript course grades was .66, with 96% of self-reported grades within a single letter grade of the transcript grade. HSGPA computed from self-reported course grades was highly correlated with transcript grade point average (r = .83). The accuracy of coursework and grades differed little by gender, race/ethnicity, and income. The results indicated that self-reported coursework and grades are reasonably accurate measures for use in education research and for preliminary screening by college admission officials.

Grade Inflation

Grade inflation is present when grades increase over time without a concomitant increase in educational achievement. A study by Ziomek and Svec (1995) examined ACT Composite scores and HSGPAs from 1990 to 1994 and found evidence for modest grade inflation. Results from a follow-up study (Woodruff & Ziomek, 2004a) covering the period of 1991 to 2003 suggested that the increase in overall HSGPA over time was largely attributable to grade inflation since the average HSGPA increase was not accompanied by a correspondingly large increase in mean ACT scores.

A more recent study by Zhang and Sanchez (2013), however, found that grade inflation has been minimal over the past decade. This study examined high school grade inflation from 2004 to 2011. Compared with the significant high school grade inflation from 1991 to 2003 (Woodruff & Ziomek, 2004a), more recent data showed no pattern of overall grade inflation or deflation across eight years. Although little evidence of overall grade inflation at US public high schools was found, significant variation across schools was identified, indicating that HSGPA inflation or deflation occurs at some high schools.

Differential Grading Standards

Another study by Woodruff and Ziomek (2004b) was designed to assess how grading standards vary across high schools. This study found that grades are more of a relative (as compared to an absolute) standard in that the interpretation of grades can vary from school to school. That is, an “A” indicates higher achievement than a “B” within a school, but an “A” at one high school does not necessarily translate to the same level of academic achievement as an “A” at another school. In addition to overall HSGPA analyses, this study evaluated differential grading standards by subject area. For further details, see the full ACT Research Report (Woodruff & Ziomek, 2004b).

Grade inflation and differential grading standards introduce additional variability into high school grades, allowing them to differ in value from year to year and school to school. In contrast, the ACT is carefully constructed to measure the same content and have the same statistical properties from year to year, and its administration does not vary from school to school. Hence, ACT scores are a useful supplement to high school grades when attempting to make valid evaluations of college readiness.

7.1.6 Statistical Relationships between End-of-Course Exams and ACT Scores

If performance on the ACT test is influenced by mastery of high school course content, one would expect that standardized measures of achievement in specific high school courses would be predictive of performance on the ACT. Moreover, the predictive relationship should be apparent even when controlling for students' levels of achievement before high school. To test this proposition, a study examined the extent to which ACT scores are predicted by scores from end-of-course exams, controlling for pre-high school academic achievement (Allen, 2015b).

The results of the study support the proposition that performance on the ACT is related to standardized measures of achievement in high school courses in the core subject areas (English, mathematics, social studies, and natural science). Thus, the study results can be used as a source of evidence for validating use of ACT scores as measures of educational achievement. The predictive weights (standardized regression coefficients) of the end-of-course exams with closer time proximity to the ACT were larger than the predictive weight of the pre-high school achievement measure (ACT Explore®) from the same subject area. While ACT Explore scores were strong predictors of ACT scores, results showed that achievement in core high school courses also had a strong relationship with ACT scores. That is, students who mastered content from core high school courses were more likely to achieve high ACT scores. In comparison to the McNeish et al. (2015) study (discussed in Section 7.1.2), the models in this study explained a greater percentage of the variation in ACT scores. Prior standardized measures of achievement predicted ACT scores better than high school course grades and courses taken, high school characteristics, noncognitive characteristics, socioeconomic status, and demographic variables. This was likely due to the standardized measures of prior achievement and end-of-course exam scores being more directly related to the outcome, which was also a standardized measure of academic achievement, relative to unstandardized variables such as high school coursework and grades.

In addition to evidence indicating that end-of-course exam performance is predictive of ACT scores, a recent study suggests that ACT test scores can also be used to help identify students who are academically prepared and may benefit from some of the more rigorous courses offered in high schools across the nation, including Advanced Placement (AP) courses (Radunzel & Allen, 2020). The study found that ACT test scores were positively related to AP exam scores and were good predictors of success. The recommended linkages to AP exam success were defined in two ways (receiving a score of 3 or higher on the AP exam or receiving a score of 4 or higher on the AP exam) and were developed in relation to content-relevant scores for most courses.

7.1.7 Statistical Relationships between Noncognitive Factors and ACT Scores

ACT has conducted substantial research examining the relationship between ACT scores and noncognitive student characteristics such as educational plans, academic behaviors, and perceptions of self. Some of these characteristics should be considered aspects of social and emotional learning, which encompasses a broad array of interpersonal, self-regulatory, and task-related behaviors important for adaptation to and successful performance in education and workplace settings (Casillas, Way, & Burrus, 2015).

When students register for the ACT, they are asked to provide information about their background, interests, needs, and plans in the Student Profile Section (SPS) of the ACT. Correlations were calculated between selected variables and ACT scores for the 2016 ACT-tested graduating class. As shown in Table 7.4, students who described their high school curriculum as college-preparatory in nature (r = .31 to .35) and aspired to higher educational levels (r = .32 to .36) tended to have higher ACT scores. Those who reported not needing help with their reading (r = .02 to .17), study skills (r = .07 to .09), or mathematics skills (r = .10 to .26) also tended to have higher ACT scores.

Table 7.4. Correlations Among ACT Scores and Background Characteristics

ACT score	College-preparatory curriculum[a]	Educational plans[b]	Does not need help in[c]		
			Reading	Study skills	Math
English	.34	.34	.14	.08	.13
Math	.32	.33	.02	.08	.26
Reading	.31	.32	.17	.07	.10
Science	.31	.32	.09	.08	.18
Composite	.35	.36	.12	.09	.18

Note. All p-values < .0001.

[a] Responses were coded 1 (college preparatory) and 0 (business or commercial, vocational-occupational, other, or general).

[b] Responses were coded 1 to 5 (vocational-technical program, associate degree, bachelor's degree, 1 to 2 years of graduate program, professional degree).

[c] Responses were coded 1 (do not need assistance) and 0 (need assistance).

A study by McNeish et al. (2015) examined the relationships between students' noncognitive characteristics, high school coursework and grades, school characteristics, demographic characteristics, and ACT scores, with an emphasis on noncognitive measures related to students' academic goals, behaviors, perceptions, and parental involvement. Consistent with earlier studies by Noble et al. (1999a, 1999b), this study's results indicated that noncognitive variables provided incremental improvement to the prediction of ACT scores; however, relative to coursework taken, grades, and school characteristics, the effect was small (i.e., less than 10%). It is important to note that students' noncognitive characteristics alone explained 29% of the variance in HSGPA, highlighting the importance of noncognitive factors to future academic achievement (for further details, see McNeish et al., 2015). Given that HSGPA entered the model first, any overlap in variance accounted for in ACT scores by HSGPA and noncognitive characteristics would be attributed to HSGPA.

In another study by Noble et al. (2006), structural equation modeling results indicated that ACT scores were directly related only to academic achievement in high school as measured by grades earned and coursework taken. Education-related accomplishments, as well as activities and perceptions of self and others (noncognitive measures), had only indirect relationships to ACT scores through academic achievement. In sum, findings from these studies suggest that noncognitive characteristics are associated with students' choices of high school coursework and the grades they earn in those courses, which, in turn, are strongly related to ACT scores.

Research has also explored the joint contributions of social-emotional learning competencies and academic achievement in earlier grades to predictions of ACT scores and Benchmark attainment. As shown in Figure 7.1, among students with similar 8–9th grade academic achievement as measured by ACT Explore, those with higher social-emotional learning scores, as measured by ACT Engage®, tended to meet more Benchmarks (Casillas, 2013). In particular, among students in the top 25% of Explore scores, students with high Engage scores met 3.41 Benchmarks, on average, compared to 2.56 for students with low Engage scores (Figure 7.1).

Social and emotional skills such as sustaining effort (how your actions demonstrate diligence, effort, organization, self-control, and compliance with rules) or maintaining composure (how your actions demonstrate relative calmness, serenity, and the ability to manage emotions effectively) are related to many positive outcomes including academic achievement (e.g., Poropat, 2009). ACT score data for 6,747 students were matched to data from Mosaic™ by ACT®—a social and emotional learning (SEL) assessment measuring five important social and emotional skills: sustaining effort, getting along with others, maintaining composure, keeping an open mind, and social connection. As shown in Figure 7.2, there was a clear positive relationship between social and emotional skills and ACT scores. For example, there was an average ACT Composite difference of 4 points when comparing the Low to High groups on sustaining effort, which is equivalent to the expected ACT Composite score increase for more than one year of instructional time.

Figure 7.1. Average Number of ACT Benchmarks Met, by Explore and Engage Grades 6–9 Academic Success Index

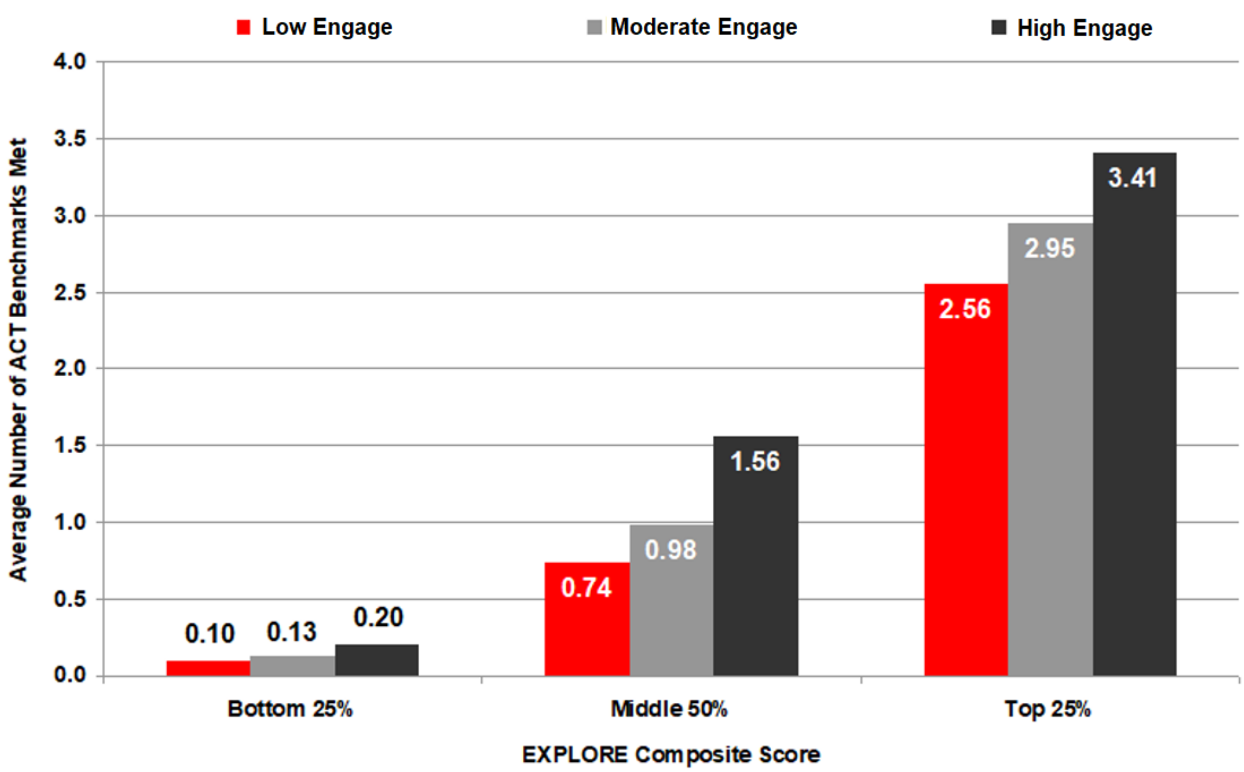
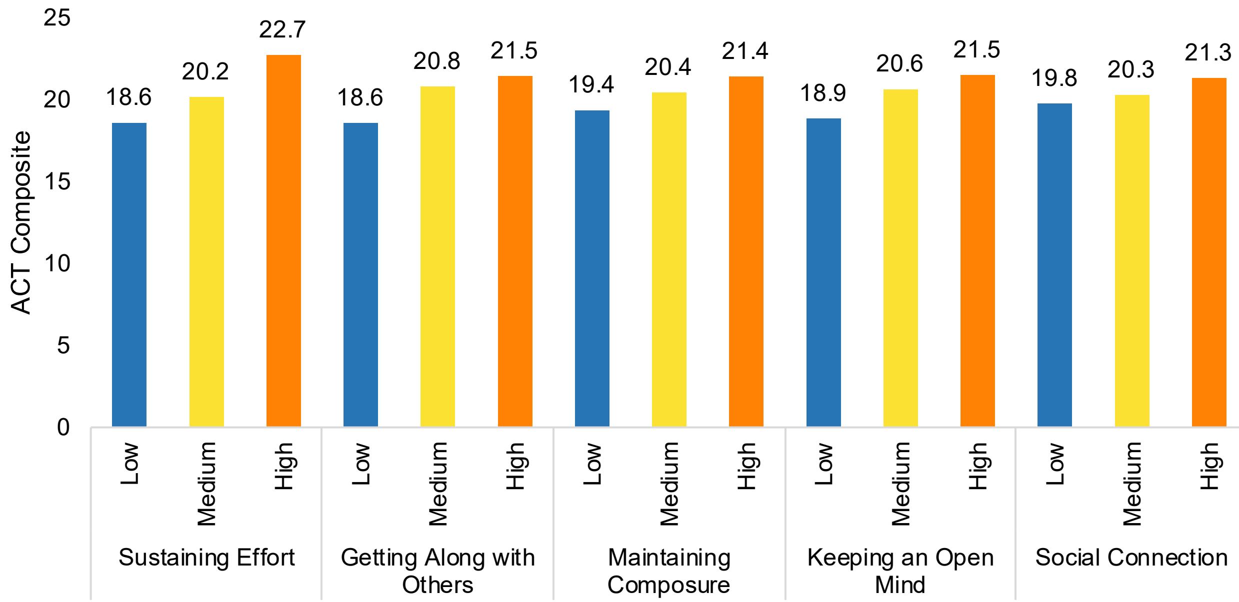


Figure 7.2. Average ACT Composite Score by Score Range on Mosaic by ACT

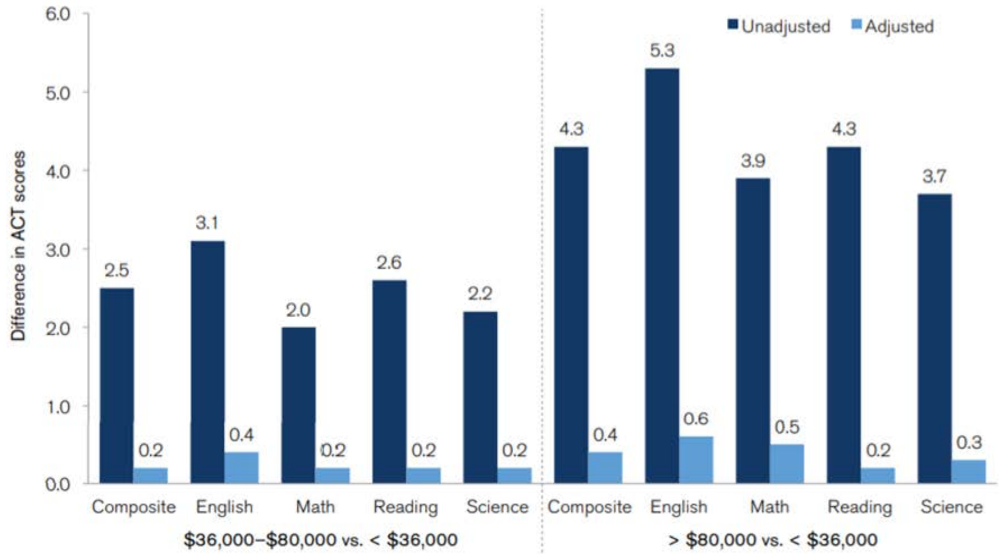


7.1.8 Understanding Group Differences on the ACT

Equity and fairness issues are important concerns of educators. Researchers have examined the strength of associations between ACT performance and predictors such as coursework, course grades, student and high school characteristics, and educational plans by race/ethnicity, gender, and/or annual family income (e.g., Noble et al., 1999a, 1999b; Noble, Crouse, Sawyer, & Gillespie, 1992; Noble & McNabb, 1989; Chambers, 1988). Their findings suggest that differential performance is largely attributable to differential academic preparation across student demographic groups.

Results from a study by McNeish et al. (2015) support the hypothesis that differential performance on the ACT results from differential academic preparation, regardless of race/ethnicity, gender, or annual family income. Specifically, after accounting for HSGPA, high school coursework, school characteristics, and other noncognitive factors, socioeconomic status and other demographic characteristics collectively accounted for a small percentage of the variance in ACT scores (4% or less). Additionally, differences in ACT scores among racial/ethnic, family income, and parental education level groups were substantially reduced when students' academic preparation levels were taken into account. School-level demographic characteristics, along with other school-level characteristics, were included in the models to account for high school attended. For example, students from high-income households (>$80,000) tended to earn ACT Composite scores that were 4.3 points higher than students from low-income households (<$36,000) (Figure 7.3). However, after accounting for student and school-level differences among the two groups, the adjusted difference in ACT Composite score was reduced to 0.4.

Figure 7.3. Unadjusted and Adjusted Mean Differences in ACT Scores by Family Income



Findings from the McNeish et al. study (2015) are consistent with results from earlier studies on this topic (Noble et al., 1999b; Schiel et al., 1996). Moreover, a more recent study that focused on deriving a high school academic rigor index (HSAR) revealed that differences in the HSAR helped to explain differences in ACT Composite scores across racial/ethnic groups (Allen, Mattern, & Ndum, 2019). HSAR was derived by optimizing the prediction of first-year college grades based on high school courses taken, grades, and indicators of advanced coursework. In the study results, the mean difference in ACT Composite score for White and African American students was reduced from 4.4 points to 2.8 points after adjusting for the HSAR index.

7.1.9 Relationship between Test Preparation and ACT Performance

The ACT assessment measures much of the knowledge and many of the skills taught in high school. Thus, it would stand to reason that long-term learning in school, rather than short-term preparation focused on test format or test-taking skills, would be the best form of preparation for the ACT. To understand better the relationship between test preparation and ACT scores, several studies were conducted to examine score increases associated with short-term test preparation activities. An analysis by Scholes and Lain (1997) suggested that preparing with practice tests for two or more hours was associated with slightly higher ACT Composite scores for first-time testers when controlling for HSGPA and grade level, but using workbooks or taking a preparation course were not. In a follow-up study, Scholes and McCoy (1998) estimated the average difference in ACT Composite scores between examinees who did and did not participate in different types of short-term preparation (workbooks and courses, workshops, or computer software) and long-term preparation (taking or planning to take a recommended core curriculum and taking or planning to take advanced courses in mathematics or science). Results showed that long-term preparation was related to much higher scores for first-time testers than short-term strategies.

In a series of three studies, Schiel and Valiga (2014a, 2014b, 2014c) investigated the association between test preparation activities and ACT performance for repeat testers. In these studies, repeat testers were surveyed about their engagement in test preparation before taking the ACT test. Regardless of preparation, the results revealed average ACT Composite scores increased slightly from the first to the second test. The observed increase, even for students who did not engage in supplemental preparation, possibly reflects a practice effect from taking the test previously or the effect of additional classroom instruction between testing occasions (Camara & Allen, 2017). In general, students who engaged in test preparation activities prior to taking the ACT for the second time exhibited a slightly greater average ACT Composite score gain compared to those who did not. A similar result was found in two follow-up studies that moved beyond descriptive statistics by statistically controlling for student background characteristics (e.g., race/ethnicity, gender, first ACT Composite score) and test-related behaviors (e.g., experienced notable stress or anxiety, had adequate sleep). First, a study by Schiel (2020) using a similar data source from a more recent cohort of students found that the adjusted average ACT Composite score gains were larger for students who reported that they prepared for the second test, compared with those who did not prepare (1.22 points vs. 0.85 points, respectively). A subsequent study employed quasi-experimental methods to allow for a causal examination of the efficacy of test preparation (Moore, Sanchez, & San Pedro, 2019). Results indicated that students who prepared for a second ACT test earned Composite scores that were 0.71 points higher on average than the scores of students who did not prepare.

Most survey respondents from the Schiel and Valiga studies (2014a, 2014b, 2014c) indicated that test preparation built their confidence, familiarized them with the test, refreshed their memory of content areas, and helped them understand subject matter. Respondents who perceived utility in test preparation for their second test exhibited greater average score increases than respondents who did not.

Further analyses from the Schiel and Valiga studies (2014a, 2014b, 2014c) revealed that test takers who spent more time engaging in test preparation activities tended to increase their scores more than those who spent less time. Other studies that have focused on the utilization of specific test preparation products have reached similar conclusions. For example, two studies by Sanchez (2019, 2020) found a positive association between greater ACT Online Prep (AOP) usage and ACT score gains, with AOP usage defined by variables such as the number of active days in AOP, the number of practice sessions completed, the number of full-length practice tests completed, the number of system resets, or the number of hours spent on AOP activities. For example, Sanchez (2020) reported that students who used AOP for less than seven hours saw greater ACT Composite score gains on average than students who did not use test preparation (1.06 vs. 0.67). In comparison, the average score gains for students who used AOP for seven or more hours were more than double those for students who did not use test preparation (1.29–1.37 vs. 0.67). Similarly, Payne and Allen (2019) found that ACT Composite score gains increased with more time spent on ACT Academy—a free, online platform that provides assessments and individualized learning plans to help students improve their academic skills. For example, compared to students who did not engage in test preparation prior to retesting with the ACT, the average ACT Composite score gain was 0.5 points higher for students who spent one hour using ACT Academy and 1.3 points higher for students who spent at least six hours using ACT Academy.

In these studies, average increases were approximately 1–2 points, depending on the duration of preparation. These studies suggest that the average effect of preparation is small, but larger score increases (and decreases) should be expected for individual students and for different student subgroups. For example, Sanchez (2018) found that, while students from all racial/ethnic groups benefitted from AOP, the average benefit to students of color (about 1 point or more) was greater than the average benefit to White students (under half a point). Results from the various studies described suggest some mechanisms to explain the relationship between test preparation and higher ACT performance such as increasing confidence, becoming familiar with test format, and refreshing or teaching subject matter.

7.1.10 Using ACT Scores for the Identification of Students with Very High Academic Achievement for Gifted and Talented Programs

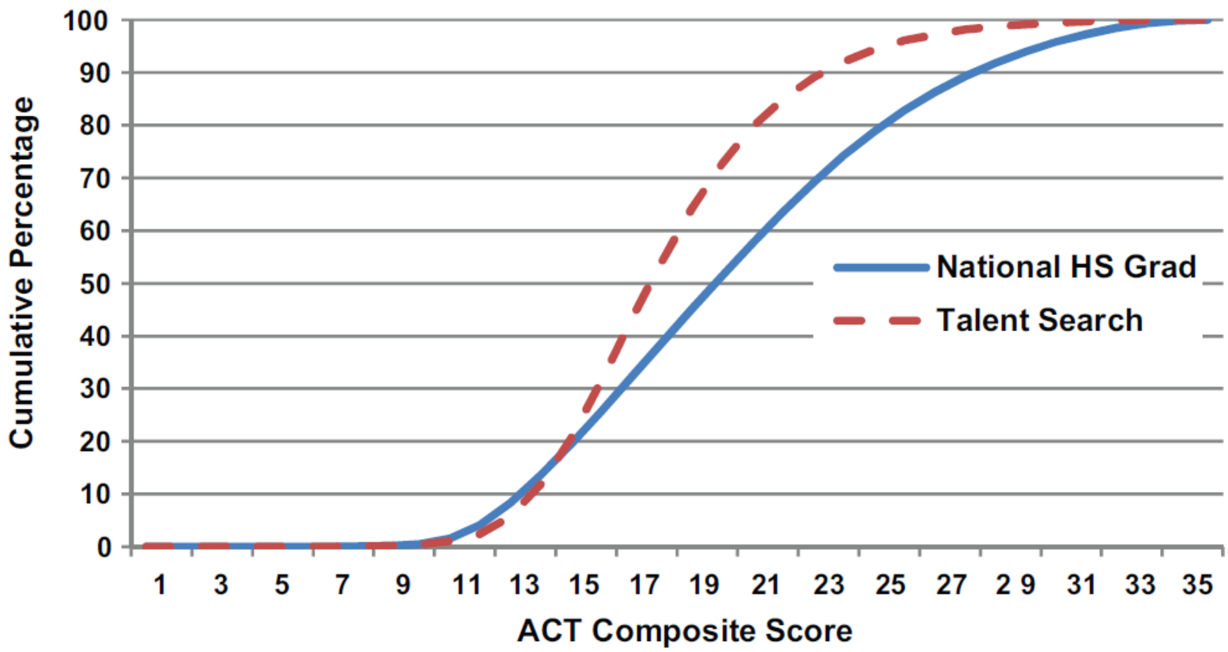
ACT scores have, over the years, been used successfully by national talent search programs to identify academically talented youth. Talent search programs provide these individuals with such services as advanced-level course materials, recognition ceremonies, and special residential programs. In a typical talent search program, 7th- or 8th-grade students who score very high (e.g., top 3%) on in-school standardized achievement tests are invited by the program to take the ACT. Those applicants earning very high ACT scores are then invited to participate in a special residential program or recognition program.

Figure 7.4 displays two ACT Composite score cumulative distributions, one representing the scores of 2016 high school graduates and the other representing the scores of a group of talent search applicants. The score distribution for the 2016 high school graduates (N = 2,090,342) in this figure was based on students who took the ACT during their sophomore, junior, or senior year and who graduated from high school in the spring of 2016. Only the most recent ACT score of each high school student was retained for analysis. The score distribution for talent search applicants was based on data from 40,562 students who took the ACT during 6th, 7th, or 8th grade in 2016 and sent their scores to a particular national talent search program.

Figure 7.4 reveals that the cumulative distribution for the 2016 ACT-tested graduating class is shifted slightly to the right of the cumulative distribution for the students identified by talent search programs who took the ACT in 6th, 7th, or 8th grade (average ACT Composite score of 20.8 vs. 18.6, respectively). This figure does not suggest either a floor effect (e.g., the assessment is too challenging for this student population and everyone scores near the bottom of the score scale) nor a ceiling effect (e.g., the assessment is too easy for this student population and everyone scores near the top of the score scale), indicating that ACT scores can meaningfully assess the educational achievement of academically talented students in 6th, 7th, and 8th grade. This is further substantiated by a study that compared ACT Composite scores earned by 7th-graders from an earlier cohort to ACT Composite scores earned by the same students four or five years later (Allen, 2016a). In that study, the average score gain was 9.4 points. This suggests a lack of a ceiling on the ACT for such students, though average score gains were lowest for students who scored near the top of the ACT Composite scale as 7th-graders.

A study by Schiel (1998) examined the academic benefits in high school of an intensive summer program for academically talented 7th-graders. The results of the study suggested that participation in summer residential programs is positively related to academically talented students' subsequent academic performance in high school. For more details, see the full ACT Research Report (Schiel, 1998).

Figure 7.4. ACT Composite Cumulative Percentages for 2016 ACT-Tested High School Graduates and Talent Search 6th-, 7th-, and 8th-Grade Students



7.1.11 Score Interpretations for Examinees Using English Learner Supports

As described in Chapter 4, some English learners (ELs) are eligible for language supports while taking the ACT. Analyses from a preliminary study suggest that providing supports and accommodations on the ACT to ELs may improve their scores and provide a more accurate reflection of their true achievement levels (Moore, Huang, Huh, Li, & Camara, 2018). Data for the study were gathered in states that administered the ACT statewide, captured students' EL status, and provided state-approved accommodations to EL students.

More recently, based on analysis of data including ELs who took the ACT using ACT-approved supports, a study indicated that the reliability of ELs' ACT scores was comparable to that of other assessments and sufficiently high that it did not, by itself, raise concerns about the validity of the scores (Moore, Li, & Lu, 2020). Data for the study included a national sample of students who took the ACT test as part of 2018 state and district testing (10,235 EL and 26,378 non-EL students). ELs who tested with (27%) and without (73%) ACT-approved supports were included in the study. In the study, classification consistency analyses revealed similar agreement rates for ELs and non-Els, and there was no evidence of differential item functioning for ELs. The latter findings suggested that item characteristics did not introduce additional bias that would have raised concerns about score validity for ELs. Prior studies on EL students revealed slight underprediction when predicting first-year college outcomes using college admissions test scores. It is hypothesized that the use of appropriate accommodations by ELs (such as those available for the ACT test) result in ACT scores that are better predictors of first-year college outcomes by reducing construct-irrelevant variance due to non-proficiency in English. Empirical studies to substantiate this hypothesis are currently planned.

7.2 Using ACT Scores to Support College Admission Decisions

Postsecondary institutions want to admit students who will be academically successful. Attending college requires a significant investment of time, money, and other resources by students and parents, as well as by the institutions; therefore, it is in their common interest that the investment succeeds. College admission therefore involves decisions made by students, counselors, and parents—all of whom may participate in selecting the institutions to which students apply—and decisions made by institutions.

Academic achievement is one important aspect of success in college, and academic preparation is one critical determinant of academic achievement. In any postsecondary academic curriculum, a certain minimum level of academic skill is required for success; beyond the minimum required level, better academic preparation usually results in greater academic success. Therefore, it is appropriate to take into account students' academic preparation when making admission decisions.

Academic success during a student's college career requires at least a minimal level of academic success in the first year. Some students experience significant academic difficulties in their first year but have satisfactory levels of achievement in subsequent years. Students whose academic difficulties in their first year cause them to leave college cannot be considered academically successful overall. Thus, the likelihood of academic success in the first year is a reasonable factor to consider when making admission decisions. Because the ACT tests measure mastery of high-school course content, which includes the academic skills needed to succeed in typical first-year college courses, they are appropriate for use in admission.

One should keep in mind that, although the ACT tests measure important academic skills needed for success in college, no test can measure all relevant academic skills. Therefore, it is advisable to supplement ACT scores with other academic information, such as courses taken and grades earned in high school, when making admission decisions.

Moreover, academic preparation is only one determinant of academic success in college—albeit an important one. Nonacademic characteristics such as motivation, interests, and goals can also influence academic success. Therefore, admission decisions should take into account students' noncognitive characteristics as well as their academic skills. The Student Profile Section and the Interest Inventory of the ACT provide information on students' background characteristics, goals, and vocational interests. As described above, measures of social-emotional learning competencies may also provide information about students' noncognitive characteristics associated with academic success in college.

Finally, there are other outcomes of postsecondary education (e.g., students' appreciation of culture, their intellectual curiosity, their ability to work with people holding differing opinions) that are not strictly academic in nature but that may be considered important educational outcomes of an institution. If an institution is able to define and defend its nonacademic goals and is able to collect information on student characteristics related to them, then such information could also be used in making admission decisions. Of course, using nonacademic characteristics to predict the achievement of nonacademic goals needs to be validated, just as using test scores to predict the achievement of academic goals must be validated.

7.2.1 Statistical Relationships between ACT Scores and First-Year College GPAs

If the ACT test measures characteristics important to success in the first year of college, and if first-year grades are reliable and valid measures of undergraduate academic performance, then there should be a statistical relationship between ACT scores and first-year grades. Therefore, a crucial aspect of any validity argument for using ACT scores in making admission decisions is the strength of the statistical relationships between the test scores and first-year grades.

Correlations as Validity Evidence

The Pearson correlation coefficient measures the strength of the linear relationship between two variables, such as college GPA and a test score. The absolute value of the correlation coefficient ranges between 0 and 1, with 0 indicating no relationship and 1 indicating a perfect linear relationship. A correlation near 0 would indicate that the relationship between test scores and GPA is too weak for the test to be used for college admission.

Measurement Error and Restriction of Range

Two factors attenuate the size of an observed correlation between ACT scores and GPA: measurement error and range restriction. Measurement error effectively places a cap on the observed correlation between two measures because the correlation between a test score and a course grade or GPA cannot exceed the square root of the product of the reliabilities of the two measures. Corrections for measurement error in test scores are not made when determining the operational validity of a test since the observed test scores are used in practice. However, corrections for measurement error in course grades or GPA permit an estimation of the validity of a predictor variable if the criterion was measured perfectly. Recent studies indicated that the estimated mean reliability of first-year GPA (FYGPA) ranges from .75 to .87 (Beatty, Walmsley, Sackett, Kuncel, & Koch, 2015; Westrick, 2017), which is lower than the estimated reliability of .94 for the ACT Composite score (see more on Reliability in Chapter 6). As an example, if the observed correlation between the ACT Composite score and FYGPA is .38, the reliability estimate for FYGPA is .81, and the reliability of ACT scores is set to 1.0 (no correction for unreliability), the validity coefficient for ACT Composite scores would increase from .38 to .42 ($.38/\sqrt{.81 \times 1} = .42$).

Range restriction in variables also reduces the correlation between predictor and criterion measures, and it is a concern in most institutional validity studies. Specifically, a correlation between test scores and college grades estimated from enrolled students whose academic skills were considered in admitting them will understate the theoretical correlation in the entire applicant population. This statistical problem exists at all postsecondary institutions whose admissions decisions take into account applicants' academic skills. On the other hand, if a college did not use test scores or other measures of applicants' academic skills in making admissions decisions, then applicants with low test scores, as well as those with high test scores, could enroll. In this situation, the correlation between the students' test scores and their grades would most likely be higher than if the college used test scores in making admissions decisions (Whitney, 1989).

With data from 50 postsecondary institutions, a recent validity study demonstrated the effects of range restriction on correlations between ACT scores and FYGPA and between HSGPA and FYGPA (Westrick, Le, Robbins, Radunzel, & Schmidt, 2015). The correction for range restriction increased the estimated mean correlation between ACT Composite and FYGPA from .38 to .51, and it increased the estimated mean correlation between HSGPA and FYGPA from .47 to .58 (Table 7.5). Note that the validity coefficients for ACT Composite score and HSGPA were somewhat variable across institutions—as indicated by the 90% credibility intervals—yet the corrected correlation between socioeconomic status and FYGPA did not vary across institutions.

Table 7.5. Meta-Analysis of Multi-Institutional Data—Correlations with FYGPA, Overall Analyses

Predictors	*k*	*N*	Mean observed *r*	SD_r	Est mean ρ	SD_ρ	95% CI	95% CrI
ACT Composite	50	169,818	.38	.07	.51	.05	.50, .53	.43, .60
HSGPA	50	150,305	.47	.05	.58	.06	.57, .60	.49, .68
SES	50	139,354	.12	.04	.24	.00	.24, .25	.24, .24

Note. k = number of institutional studies; SD_r = standard deviation of observed correlations; Est = estimated; SD_ρ = standard deviation of correlations corrected for artifacts; CI = confidence interval; CrI = credibility interval. Table adapted from Westrick et al. (2015).

A follow-up study demonstrated that ACT Composite scores provided incremental improvement to the prediction of FYGPA after accounting for HSGPA (Westrick, Schmidt, Le, Robbins, & Radunzel, 2021). Such studies support the use of ACT scores in combination with other measures to help inform college admission decisions. In that study, after correcting for range restriction and measurement error in both the predictors and the criterion, the estimated mean multiple correlation for the FYGPA prediction model that included ACT Composite score and HSGPA was .662, compared to .627 for the model that only included HSGPA. The incremental validity finding has been corroborated by other studies, including a study by Allen et al. (2019), which reported an increase in the multiple correlation when ACT Composite score was added to the FYGPA prediction model that included HSGPA (from .48 to .52).

Multiple Scores and Superscoring

Another factor that can impact validity coefficients is the choice of which ACT scores to include in the analysis for students who take the ACT more than once. Postsecondary institutions report using a variety of composite scoring methods for applicants with multiple test records, including averaging and taking the maximum subtest score across test occasions ("superscoring"). A recent study evaluated the validity of using ACT scores obtained from various scoring methods across test administrations (average, highest, most recent, and superscoring) for identifying students who are likely to be successful in their first year of college (Mattern, Radunzel, Bertling, & Ho, 2017). Compared to the other options, the study found that superscores were as predictive—if not more predictive—of first-year grade point average and resulted in the least amount of differential prediction when statistically controlling for the number of times tested. Specifically, the differential prediction results suggested that first-year grades for examinees who tested more often tended to be underpredicted by ACT scores. That is, retesters performed better in college than what was expected based on their test scores. However, this prediction error was lower when superscores were used instead of the other scoring methods.

A follow-up study was conducted to examine the relationship between different scoring methods and another important measure of college success: degree completion (Radunzel & Mattern, 2020). Consistent with the earlier study, this study indicated that superscores were as predictive of completing a college degree in a timely manner compared to the other scoring methods. The study also revealed that the likelihood of completing a degree for students who tested more often was underpredicted, but that superscoring resulted in the least prediction error when prediction accuracy was examined by the number of times students tested. These findings held for both the two- and four-year institution samples. These two studies provide evidence supporting ACT's new practice of reporting the ACT Superscore, since they suggest that selecting students' best scores from any test attempt results in the most predictive indicator of a student's preparedness for future academic success.

Decision-Based Statistics as Validity Evidence

The correlation coefficient is used more often than any other statistic to summarize the results of predictive validity studies. As an index of the strength of the linear relationship between first-year college grades or GPAs and admission or placement measures, a correlation coefficient can lend credibility to a validity argument. However, it does not directly measure the degree to which admission or placement measures correctly identify students who are academically prepared for college coursework. Rather, the correlation coefficient reflects the accuracy of prediction across all values of the predictor variables. Of greater interest to educators who must evaluate admission or placement systems is the correctness of the decisions made about individual students and their estimated chances of success. In this section, logistic regression and decision-based statistics are described as alternative methods for summarizing the results of predictive validity studies. Studies presented in this section demonstrate the use of this method for making admission and course placement decisions.

Suppose "success" in the first year of college can be defined in terms of some measurement that is obtainable for each student; for example, success might be defined as a student completing the first year with a GPA of C or higher in a common subset of first-year courses. Then, there are four possible outcomes of the admission decision for a particular student:

A. True positive: the student is permitted to enroll in the college and is successful there. (Correct decision)
B. False positive: the student is permitted to enroll in the college and is not successful there. (Incorrect decision)
C. True negative: the student is not permitted to enroll in the college and would not have succeeded if he or she had enrolled. (Correct decision)
D. False negative: the student is not permitted to enroll in the college and would have succeeded if he or she had enrolled. (Incorrect decision)

The sum of the proportions of students associated with outcomes A and C is the proportion of correct admissions decisions.

Note that outcomes A and B can be directly observed in existing admission systems, but outcomes C and D cannot. In principle, the proportions associated with all four outcomes could be estimated by collecting admission measures (e.g., admission test scores) on every student, while permitting everyone to enroll in the college, regardless of test score. Some of these students would be successful in the college and others would not; the relationship between the probability of success and the admission measures could then be modeled using statistical methods. From the estimated conditional probabilities of success for given values of the admission measures, estimates of the probabilities of the outcomes A–D could be calculated.

In most institutions, of course, this kind of experimentation is not done because students with low probabilities of success are generally not admitted or do not select the college. Therefore, first-year outcomes are not available for these students, and the relationship between their probability of success and their admission measures must be estimated by extrapolating relationships estimated from the data of students who actually enrolled in the college. The assumption being made is that the conditional probability of success given the selection variable(s) is the same for the nonenrolled applicants as for the enrolled students. This assumption is analogous to that for the traditional adjustment of correlations for restriction of range, which requires that the applicant and enrolled student groups have the same conditional mean and variance functions (e.g., Lord & Novick, 1968). Research at ACT has shown that accurate extrapolations can usually be made from moderately truncated data (Houston, 1993; Schiel & Harmston, 2000; Schiel & Noble, 1992).

It is possible to relate a correlation coefficient to the conditional probability of success function, but a number of strong statistical assumptions are required. A more straightforward way to estimate the probability of success is to dispense with correlation coefficients altogether and to model it directly. For example, one could use the logistic regression model

$$\hat{P}[W = 1|X = x] = \frac{1}{1+e^{-\hat{a}-\hat{b}x}} \quad (1)$$

where W = 1, if a student is successful in college;

W = 0, if a student is not successful in college; and

X is the student's admission test score.

An example of an estimated logistic function is the curve labeled "Probability of C or higher" in Figure 7.5. Note that the probability of C or higher ranges from .05 to .99, depending on the test score. Note that this curve is S-shaped and that its maximum slope occurs at the test score of 20. In logistic regression, the point at which the maximum slope occurs is called the "inflection point," and the slope of the curve at this point is proportional to the coefficient $\hat{b}$ in Equation (1). Therefore, a larger $\hat{b}$ value corresponds to a steeper slope, which indicates that the test is better at discriminating between students who will and will not succeed (for students with test scores near the inflection point).

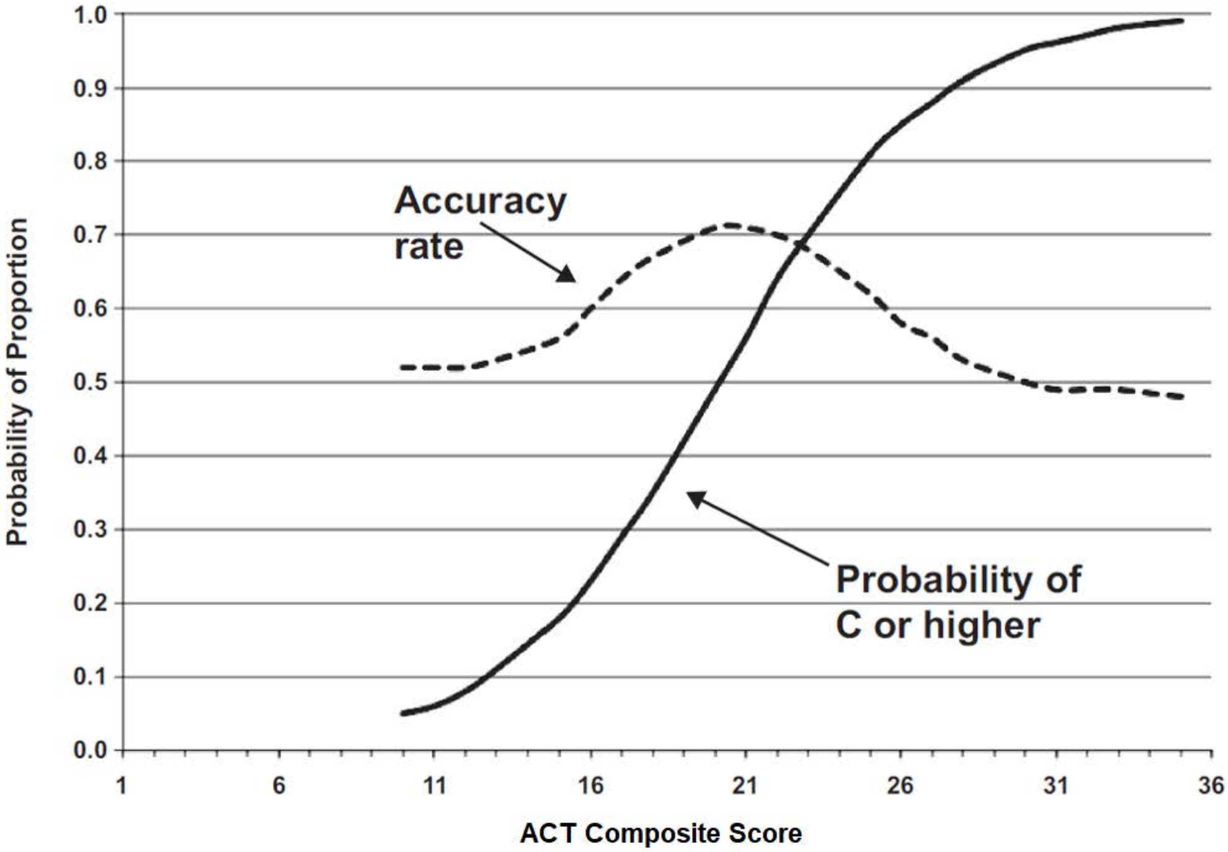
The estimated $\hat{a}$ and $\hat{b}$ coefficients in Equation (1) can be calculated by iterative least squares procedures. Given the previous discussion, the coefficient $\hat{b}$ should be positive and significantly different from zero. A coefficient near zero would result in a flat curve for the conditional probability of success.

Once estimates $\hat{a}$ and $\hat{b}$ have been obtained, estimated probabilities for the four outcomes can be calculated easily. For example, if 16 is the cutoff score on X for being admitted to an institution, then the probability of a true positive (Outcome A) can be estimated by

$$\hat{P}[A] = \frac{\sum_{x \geq 16} \hat{P}[W=1|X=x]n(x)}{N} \quad (2)$$

where $\hat{P}[W = 1|X = x]$ is Equation (1) calculated from the estimates $\hat{a}$ and $\hat{b}$, *n(x)* is the number of students whose test score is equal to x, and *N* is the total number of students in the sample. At institutions with existing admission systems, the conditional probabilities $\hat{P}[W = 1|X = x]$ in Equation (1) are calculated from data for students who enrolled in the institution. The probability $\hat{P}[A]$ in Equation (2), however, is calculated from the test scores of all students, both those who were admitted and those who were not admitted. The probabilities for outcomes B, C, and D can be estimated in a similar way.

Figure 7.5. Probability of C or Higher FYGPA and Accuracy Rate



It should be noted that admission decisions are usually made on the basis of several measures. For the purpose of illustrating how the accuracy of admission decisions can be estimated, the example uses a simplified model based on a cutoff score on a single admissions test. Students scoring at or above the cutoff score would be admitted; students scoring below the cutoff score would not be admitted. ACT does not advocate making admission decisions solely on the basis of a single measure; this example is for illustration only. Results provided later in this section illustrate how the logistic regression model may be generalized to multiple measures.

Once the estimates $\hat{P}[A]$ and $\hat{P}[C]$ are obtained, the percentage of correct admission decisions ("accuracy rate") is estimated as $\hat{P}[A] + \hat{P}[C]$, multiplied by 100. An illustration of estimated accuracy rates for different test scores is given in Figure 7.5 as a proportion. Note that the maximum accuracy rate (0.71) occurs at the inflection point in the graph of the probability of success (i.e., near a score of 20). This score is referred to as the optimal cutoff score, the score that maximizes the percentage of correct admission decisions.

The accuracy rate value corresponding to the lowest obtained test score represents the overall percentage of students who would succeed in college without using the test for admission. The difference ("increase in accuracy rate") between the maximum accuracy rate and the accuracy rate for the lowest test score is an indicator of the effectiveness of the test for making admission decisions. This statistic shows the increment in the percentage of correct admission decisions due to the use of the test. Large increases in accuracy rate correspond to a greater contribution by the test in increasing the rate of correct admission decisions. Note a selection variable has incremental accuracy if and only if its probability-of-success curve crosses 0.5 somewhere.

The ratio of true positives, $\hat{P}[A]$, to the sum of true positives and false positives, $\hat{P}[A] + \hat{P}[B]$, multiplied by 100, indicates the estimated percentage of students who would be successful of those who would be admitted using particular admission criteria. This ratio is called the "success rate." Like the probability of success, the success rate should increase as scores on the admission measure increase. The incremental success rate associated with a selection variable is the difference between the success rate associated with admitting applicants at or above the specific cutoff score and the base success rate for the lowest test score (i.e., the success rate associated with admitting all applicants).

College Admission Validity Evidence Using Decision-Based Statistics

A majority of postsecondary institutions use standardized test scores in combination with high school grades or rank for making admission and course placement decisions (Clinedinst, 2015). This activity is supported by research demonstrating the validity of using multiple measures for making college admission and placement decisions (e.g., Noble, Crouse, & Schulz, 1995; Noble & Sawyer, 2002) and the content perspective that no test can measure all the skills and knowledge needed for success in college. Using multiple measures can increase content coverage and, as a consequence, increase the accuracy of admission decisions versus using test scores alone.

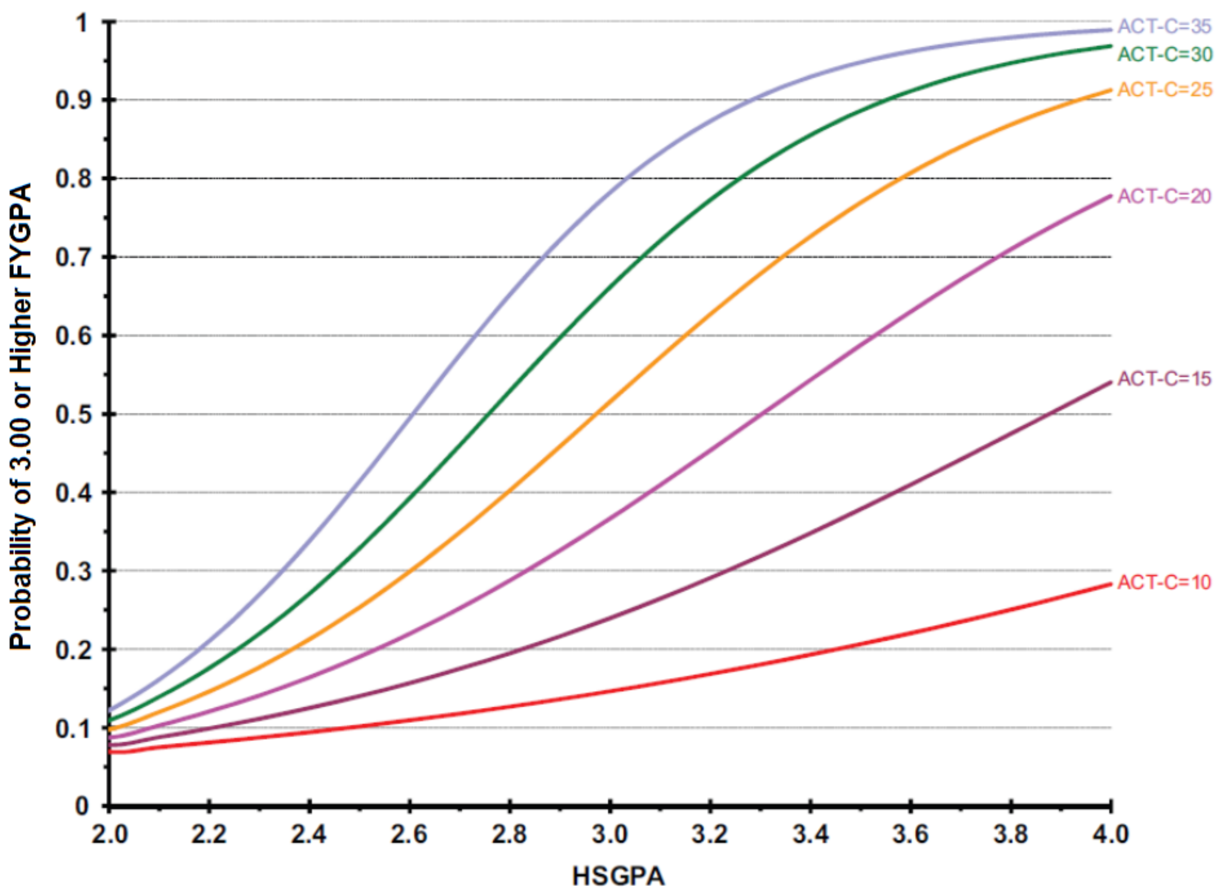
The usefulness of a selection variable for admission to college depends in large part on its predictive power, but it also depends on admission officers' goals, which are aligned to their institutions' larger goals to educate students successfully. Usefulness also depends on other issues, such as applicant self-selection and institution selectivity. To gauge the usefulness of a selection variable, one must specify the goals of using that variable. Two common goals related to academic achievement are

- maximize academic success among enrolled students and
- accurately identify applicants who would be academically successful at the institution and enroll as many of them as possible.

These goals may seem similar, but they are not identical. The first goal is related to the proportion of applicants who would succeed academically if they enrolled (i.e., the success rate). The second goal is related to the proportion of applicants that an institution correctly identifies as likely to succeed or likely to fail (i.e., the accuracy rate). Both goals, however, pertain only to institutions with some degree of selectivity in their admission policies rather than to institutions with open-admission policies.

A study was conducted to evaluate the usefulness of ACT Composite score and HSGPA for college admission decisions using the decision-based statistics discussed in the previous section (Sawyer, 2010). Using data from 192 four-year postsecondary institutions, the study evaluated whether using ACT Composite score for selection increased the success rate and accuracy rate over what would result if the institution did not use ACT Composite score. For example, Figure 7.6 shows the probability of earning a FYGPA of 3.0 or higher, given different values of HSGPA and ACT Composite score. The model coefficients for the main effects (HSGPA and ACT Composite score) and their interaction were statistically significant ($p < .001$). The interaction indicated that the slope of the probability-of-success curve increased as ACT Composite score increased. One interpretation of the interaction term is that HSGPA is more predictive among students with higher ACT Composite scores than for students with lower ACT Composite scores. This figure also illustrates that the probability of earning a FYGPA of 3.0 or higher varies dramatically among students with the same HSGPA but different ACT Composite scores. Among students with a 4.0 HSGPA, students with an ACT Composite score of 15 have a probability of .54, compared to over a .95 probability for students with an ACT Composite score of 30. Even in less extreme cases, the results illustrate that ACT Composite scores meaningfully discriminate among students with the same HSGPA in terms of which students are more or less likely to succeed in the first year of postsecondary studies.

Figure 7.6. Probabilities of Earning a 3.0 or Higher FYGPA and Being Retained Through the First Year Based on HSGPA and ACT Composite Score



Results from the Sawyer (2010) study were consistent with those from an earlier study by Noble and Sawyer (2002). Results from both studies suggest that HSGPA by itself is better than ACT Composite score by itself for some, but not for all, degrees of selectivity and definitions of success. The ACT Composite score is more useful than HSGPA in some situations—for example, when an institution is interested in high levels of success. In most scenarios, using both high school grades and test scores jointly is better than using either by itself. When using both variables, it is important to take into account the HSGPA by ACT Composite score interaction effect, as well as the main effects.

Postsecondary institutions seek high achievement for their students and want to admit students who have a good chance of being successful in college. The results from these studies suggest that ACT Composite scores provide differentiation across levels of achievement in terms of students' probable success during their first year in college.

7.2.2 Differential Prediction in First-Year College GPA Among Student Groups

Differential prediction occurs when students who have the same test scores, but belong to different population groups, have different probabilities of success. One of the effects of differential prediction is that, if an institution used cutoff scores based on students' probability of success to make admission decisions, different observed success rates could result for different population groups. For example, predictive correlations could differ among the groups. Another possibility could be that the proportion of admitted applicants who are successful (success rate) and the proportion of correct admission decisions (accuracy rate) could differ. Any such differences may result from differential validity.

Differential Prediction by Race/Ethnicity, Gender, and Family Income

A study by Sanchez (2013) investigated differential effects on student groups using ACT Composite score and HSGPA for making admission decisions. Student characteristics included race/ethnicity, gender, and income. For each student group, Sanchez examined the effect of using a total group cut score for ACT Composite score, HSGPA, or both on predicting first-year college grade point average (FYGPA).

Results indicated that the probability of earning a 2.5 or 3.0 FYGPA increased as ACT Composite score or HSGPA increased, and this was true for White, African American, and Hispanic students (Figures 7.7 and 7.8). For the two FYGPA levels (2.5 and 3.0), White students had higher estimated probabilities of success than African American and Hispanic students over most of the ACT Composite score and HSGPA scales, and Hispanic students tended to have higher estimated chances of success than African American students. When differential prediction was apparent, it tended to be of greater magnitude when HSGPA was used as the academic predictor than when ACT Composite score was used (Figure 7.8 vs. Figure 7.7). This was particularly notable for African American students scoring above a HSGPA of about 3.0. This suggested a total-group HSGPA model considerably overestimates the chances of success for African American and Hispanic students with a high HSGPA.

Figure 7.7. Estimated Probabilities of Achieving Specific FYGPA Levels Based on ACT Composite Score, by Race/Ethnicity

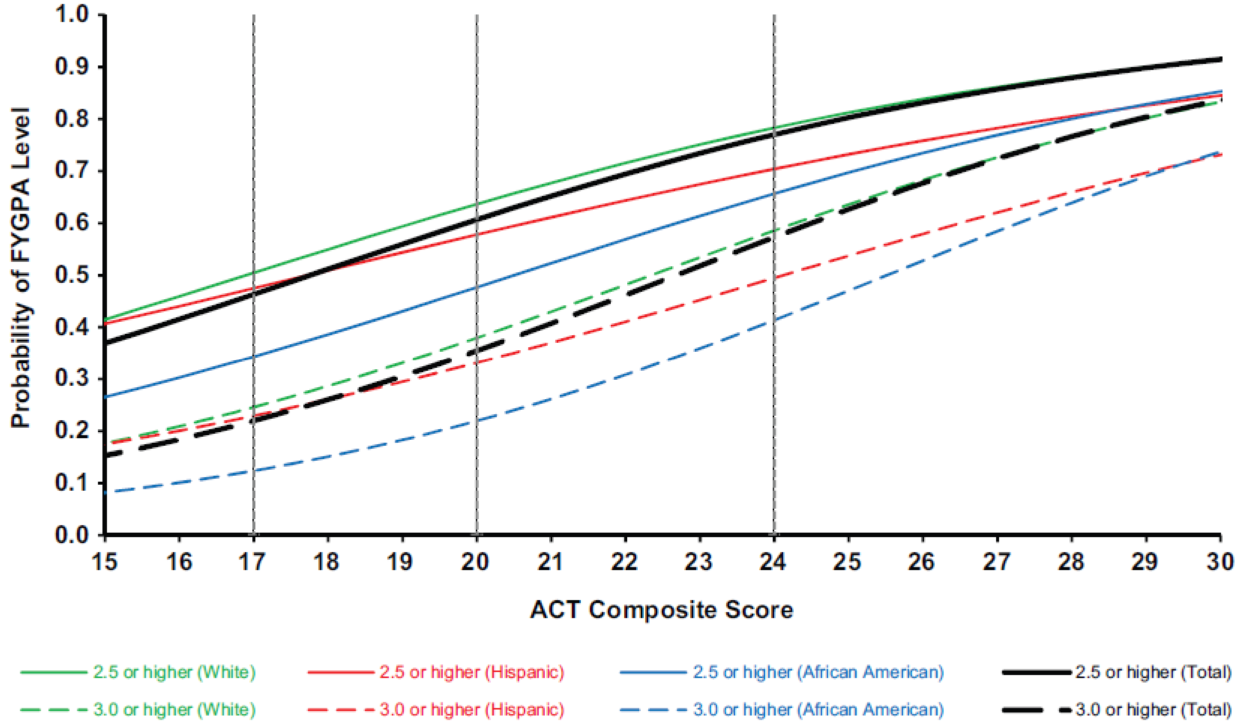
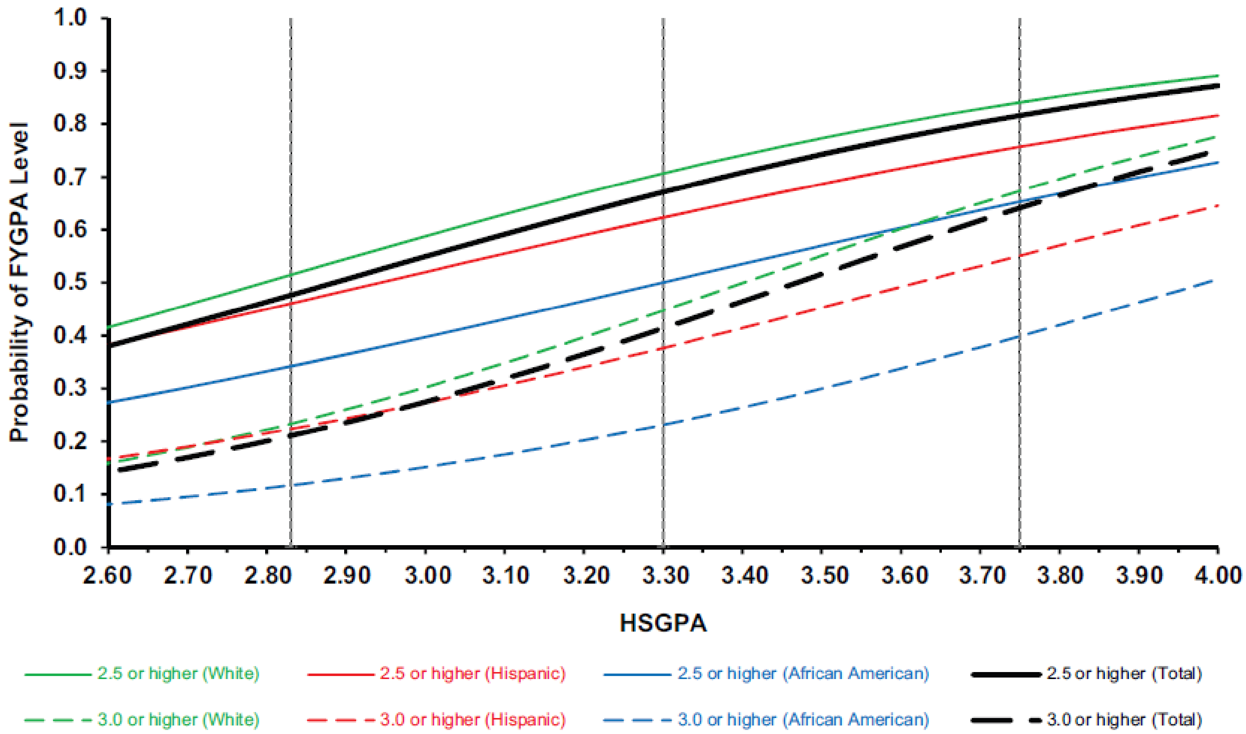


Figure 7.8. Estimated Probabilities of Achieving Specific FYGPA Levels Based on HSGPA, by Race/Ethnicity



Consistent with Figures 7.7 and 7.8, the median probabilities of success across institutions based on a total-group cutoff for racial/ethnic groups tended to show a pattern of underprediction for White students and overprediction for both Hispanic and African American students (Table 7.6). The joint ACT Composite score and HSGPA model tended to produce the most favorable accuracy rates and success rates, on average, across the racial/ethnic groups. For the 3.0 or higher FYGPA success level, median accuracy rates were highest for African American students and lowest for White students. Moreover, the increase in accuracy rates associated with using ACT Composite score and HSGPA jointly as predictors was greater for African American and Hispanic students than for White students.

Table 7.6. Median Statistics for Predicting 3.0 FYGPA or Higher by Race/Ethnicity Across 236 Institutions

Predictor	Total-group cutoff	Race/ ethnicity	Median			
			Probability of success	Maximum accuracy rate	Accuracy rate increase	Success rate
ACT Composite score	23	White	.54	71	25	68
		African American	.36	86	71	46
		Hispanic	.45	78	56	53
HSGPA	3.4	White	.52	72	22	68
		African American	.27	81	64	37
		Hispanic	.42	75	49	52
ACT Composite score & HSGPA	–	White	.51	75	30	70
		African American	.32	87	73	48
		Hispanic	.43	81	61	55

Note. Multiple combinations of ACT Composite score and HSGPA correspond to a .50 probability of success in the joint models.

Overall, the study revealed that, across student groups, the joint use of ACT Composite score and HSGPA resulted in greater prediction accuracy than either predictor alone. Furthermore, the use of a total-group cutoff score for both ACT Composite score and HSGPA, individually and in combination, slightly overpredicted the probability of success of Hispanic and African American students, males, and students from low-income households. Both ACT Composite score and HSGPA, individually and in combination, slightly underpredict the probability of success of White students, females, and students from higher-income households. These findings suggest, therefore, that students who are African American, Hispanic, or from low-income households are not disadvantaged when test scores, alone or in combination with other predictors, are used to predict future performance in college and make admission decisions.

These results are further corroborated by findings from a parallel study (Radunzel & Noble, 2013) that examined the differential effects on student demographic groups of using ACT scores and HSGPA for predicting long-term college success through degree completion. In general, research indicates that the use of multiple measures helps capture a more holistic view of student readiness and improves prediction of postsecondary success. As a case in point, results from a study by Mattern, Sanchez, and Ndum (2017) suggested that including noncognitive measures such as academic discipline (the amount of effort a student puts into schoolwork and the degree to which a student sees himself or herself as hardworking and conscientious) into a prediction model that already included ACT Composite score and HSGPA increased the variance accounted for in FYGPA and reduced the amount of differential prediction by gender.

Differential Prediction for Students Testing with Accommodations

Since the enactment of the Individuals with Disabilities Education Act (IDEA) in 1975, the total percentage of students enrolled in public schools with disabilities has increased from 8.3% (1976–1977) to 8.8% (2004–2005), and the percentages have hovered around 13% to 14% from 2005 to 2019 (Snyder & Dillow, 2013; National Center for Education Statistics, 2021). The number of students who elect to take the ACT under special conditions continues to grow. Accommodations for eligible students with disabilities are discussed in Chapter 4. Briefly these include but are not limited to the following:

- large type edition,
- Braille edition,
- DVDs edition, and
- reader's script administration.

Table 7.7 shows average ACT scores for a sample of 436,695 students who tested with accommodations and 7,252,520 students tested without accommodations between September 2016 and July 2020. On average, accommodated students achieved lower ACT scores than non-accommodated students, with accommodated students earning an average ACT Composite score of 18.4 and non-accommodated students earning an average ACT Composite score of 20.5. However, there were large differences in performance by disability category.

Because of the growing number of students with disabilities, it is important to demonstrate that a student's ACT scores and HSGPA are valid predictors for college success, not only for students tested under regular conditions but also for students with disabilities who received testing accommodations. Several prior studies have demonstrated the validity of ACT Composite score and HSGPA in predicting the FYGPA of students with disabilities who received a testing accommodation (Laing & Farmer, 1984; Ziomek & Andrews, 1996). A more recent study by Huh and Huang (2016) examined this issue.

Huh and Huang (2016) found that ACT tests scores obtained under accommodations for students with disabilities are predictive of FYGPA. Moreover, using multiple measures provides a more accurate prediction of special-tested students' chances of succeeding in college. Specifically, this study found that a prediction model that uses both ACT Composite score and HSGPA is a good model to predict actual college FYGPA for both students testing with accommodations and those testing without accommodations. Full results can be found in ACT Research Report 2016-7.

Table 7.7. Average ACT Scores for Accommodated and Non-Accommodated Examinees, by Disability Category

Disability category	*N*	Percent	English	Math	Reading	Science	Composite
Neurodevelopmental disability	**317,603**	**73%**	**16.0**	**17.4**	**18.0**	**17.8**	**17.4**
ADHD	87,626	28%	18.8	19.3	20.7	19.8	19.8
Autism	18,076	6%	18.5	18.6	19.8	19.5	19.3
Communication disorder	6,954	2%	14.6	16.5	16.4	16.7	16.2
Intellectual disorder	9,179	3%	11.2	14.3	13.1	14.0	13.3
Learning disability, math	16,286	5%	14.3	15.0	16.3	15.3	15.4
Learning disability, reading	44,479	14%	15.4	17.7	17.9	18.0	17.4
Learning disability, writing	3,829	1%	15.4	17.9	17.6	18.1	17.4
Learning disability, unspecified	1,722	1%	13.8	15.8	15.9	15.9	15.5
Motor disability	431	0%	21.4	21.1	22.3	21.5	21.7
Other neurodevelopmental disability	71	0%	26.1	23.7	26.7	24.9	25.5
Multiple neurodevelopmental disabilities	128,950	41%	14.6	16.5	16.8	16.8	16.3
Physical/sensory disability	**32,001**	**7%**	**19.5**	**19.9**	**21.3**	**20.5**	**20.4**
Hearing impairment	4,738	15%	15.1	17.1	17.4	17.7	16.9
Motor impairment	2,014	6%	20.2	19.9	22.6	21.0	21.0
Visual impairment	4,777	15%	20.1	20.4	22.5	21.0	21.1
Other physical/sensory disability	19,139	60%	20.3	20.5	21.7	21.0	21.0
Multiple physical/sensory disabilities	1,333	4%	19.8	20.0	22.1	20.9	20.8
Psychological disability	**28,712**	**7%**	**20.3**	**19.8**	**22.3**	**20.6**	**20.9**
Anxiety	10,261	36%	23.3	21.9	25.4	23.1	23.6
Conduct disorder	9,899	34%	14.9	16.0	16.7	16.3	16.1
Depression	1,000	3%	21.6	20.7	23.8	21.6	22.0
Other psychological disability	1,089	4%	21.6	21.2	23.3	21.9	22.1
Multiple psychological disabilities	6,463	23%	23.4	21.8	25.5	23.0	23.5
Multiple disabilities	**54,416**	**12%**	**20.7**	**20.2**	**22.6**	**21.2**	**21.3**
Other disability	**3,963**	**1%**	**19.1**	**19.6**	**20.8**	**20.1**	**20.0**
All accommodated examinees	436,695	–	17.1	18.1	19.2	18.6	18.4
Non-accommodated	7,252,520	–	19.7	20.2	21.0	20.5	20.5

Note. Left-aligned percentages are the percentages of all accommodated examinees, and right-aligned percentages are the percentages of examinees within that high-level disability category.

7.3 Using ACT Scores to Support Course Placement Decisions

One common use of the ACT tests is to facilitate placement into first-year college courses. This section summarizes research conducted on the effectiveness of ACT scores for this use.

At many postsecondary institutions, there are two levels of first-year courses: "standard" courses in which most students enroll and "remedial" or "developmental" courses for students who are not academically prepared for standard courses. At some institutions, there may also be "advanced" or "honors" courses for exceptionally well-prepared students.

In all cases, one can think of placement as a decision about whether to recommend that a student enroll in an “upper-level” or a “lower-level” course. The names “upper-level” and “lower-level” may refer variously to standard and remedial or developmental courses, or to advanced and standard courses. Placement systems typically identify students with a small chance of succeeding in an upper-level course and therefore recommend that they enroll in a lower-level course.

7.3.1 Placement Validity Argument Based on ACT Content

A validity argument for a placement test can, in part, be based on subject matter content. The full ACT test is intended to measure academic skills and knowledge that are acquired in typical college-preparatory curricula in high school and that are essential for academic success in the first year of college. The content specifications of the ACT are based on the recommendations of nationally representative panels of secondary and postsecondary educators (see Chapter 2). Determining the content alignment between ACT tests and a particular course at a given postsecondary institution must, of course, be done by faculty at the institution who know the course content. ACT therefore recommends that faculty and staff review the ACT test content and specifications to determine their relationship to the first-year curriculum as a preliminary step in deciding whether to use the ACT for first-year course placement.

Given that the content of the ACT tests is related to the skills and knowledge required for success in college, and given that course grades are reliable and valid measures of educational performance in the course, there should be a statistical relationship between test scores and course grades. If there is close content alignment between an ACT test and the college course, then it is reasonable to expect that students with higher ACT scores will tend to be more successful in the college course than students with lower ACT test scores. If this expectation of ACT scores is borne out in empirical studies, then it is appropriate to consider using the tests for course placement.

As noted previously, the ACT does not measure all aspects of students’ readiness for all first-year college courses. Therefore, it is advisable to consider using additional measures such as high school coursework and grades, scores on locally developed placement tests, or noncognitive measures, in addition to ACT scores when making placement decisions. Feasibility and cost are two key issues in deciding whether and how to use additional measures of academic skills for course placement.

7.3.2 Statistical Relationships between ACT Scores and Course Grades

ACT has collected course grades from postsecondary institutions specifically to examine the effectiveness of the ACT tests for placement. Results from these analyses provide validity evidence for using ACT scores for placement.

Data and Method

Grade data were gathered from entry-level courses at two-year and four-year institutions and included several different course types. Within each institution, courses that had at least 50 students who earned a course grade and had an ACT score on the corresponding section test were included in the analyses. The sample for each course was weighted to match the population of ACT-tested enrollees at each institution in terms of gender, race/ethnicity, ACT Composite score level, and HSGPA level.

Logistic regression models were used to estimate probabilities of success for each course for each institution (data permitting). Course success, which was defined as earning a grade of B or higher, was predicted from the relevant ACT score. Only courses with success rates between 20% and 80% and with logistic regression curves that crossed the .50 probability level were retained in the analyses.

At each ACT score, the success and accuracy rates were estimated from the probabilities of success obtained from the logistic regression model (see section 7.2.1 for descriptions of these statistics). These decision-based statistics were then summarized across institutions by course type. To assess validity, accuracy rates were summarized at the institution-specific optimal cutoff score, which is the ACT cutoff score that, if used for course placement, would provide the most accurate prediction of course success. When examined across a range of possible cutoff scores for a given institution, the accuracy rate typically peaks at a specific score and then decreases as the score increases.

This maximum value, which corresponds to a .50 probability of success, is the “optimal” cutoff score for a given course. There are four reasons why success was defined as a grade of B or higher rather than C or higher:

1. The statistical model would be unstable if success or failure occurs rarely, and grades below C are fairly uncommon in most courses.
2. If the optimal cutoff score is used for course placement, the least-qualified student allowed into the course has about a 50% chance of being unsuccessful. If success is defined as a grade of C or higher, that means the least-qualified student has about a 50% chance of getting a grade of D or F. It would seem poor policy to place a student into a course with that high of a chance of needing to repeat the course due to poor grades.
3. The success criterion of B or higher results in grade distributions that more closely follow those currently found in colleges. As noted above, grades below C are fairly uncommon in most courses. Moreover, the mean FYGPA tends to be closer to 3.0 than to 2.0 in recent studies (Allen & Radunzel, 2016; Radunzel & Noble, 2012; Sawyer, 2013a).
4. Prior studies have shown that students who earn B or higher grades in the first year of college are much more likely to earn a college degree, relative to those who earn lower grades (Allen, 2013).

Validity can also be examined by the strength of relationship between the predictor (ACT scores) and course success. The logistic regression model is defined by intercept and slope coefficients, and the slope indicates the strength of the relationship. To summarize the strength of the relationship, median logistic regression slopes are also provided.

Results

Accuracy Rate

Table 7.8 provides the summarized results for 17 courses. For all courses, the median accuracy rate at the optimal cutoff score was at least 62% (refer back to section 7.2.1 for an overview of decision-based statistics). Thus, a typical institution using the ACT optimal cutoff score from their data could expect that 62% or more of the placement decisions made would be correct. Differentiating by course type shows that Intermediate Algebra courses (using the ACT mathematics score for placement) was among the courses with the lowest median accuracy rate (62%) and Composition II courses (using the ACT English score for placement) had the highest (68%). Although the magnitude of the accuracy rates might be used as evidence of placement validity, one needs to compare the maximum accuracy rate at the optimal cutoff score to the accuracy rate that would result without placement—the accuracy rate that would result if all students were allowed to enroll in the course. The difference between these two values for each course represents the increase in the accuracy rate resulting from using ACT test scores for placement. For example, for College Algebra the median accuracy rate was 66%, and the median increase in accuracy rate was 13 percentage points. Thus, if all students were allowed into the course, the expected accuracy rate would be 53%.

Compared with English courses, mathematics, social science, and natural science courses tended to show higher increases in accuracy rates when using ACT test scores for placement. For English courses with sufficiently large samples, the course placement statistics were assessed for ACT English scores. English courses tend to have higher percentages of students earning a B or higher, so the accuracy rates are well above 50% without using any placement measures. This leads to smaller increases in accuracy rates regardless of the placement variable (e.g., standardized tests, high school grades, locally developed placement tests, or performance assessments).

Success Rate

The median success rates at the optimal cutoff score ranged from 60% in Economics and Intermediate Algebra courses to 68% in the Composition courses. This suggests that an institution using its optimal ACT cutoff score typically could expect that at least 60% of the students who were placed in the standard course would obtain a grade of B or higher.

The median logistic regression slopes measure the strength of relationship between ACT test scores and the course success outcomes. Specifically, the slopes represent the change in the log-odds of success for each one-point increase in the test score. For example, the log-odds of success in Biology increased by 0.196 for each one-point increase in the ACT science score. Consistent with prior studies (Allen, 2013), the slopes tended to be larger for mathematics and natural science courses than for English and social science courses.

The optimal cutoff score for a given course varies across institutions (Allen, 2013). Variation in grading standards and course difficulty across institutions can contribute to this variation in optimal cut scores. Because results vary across institutions, institutions should collect their own course outcome data and determine their placement cutoff scores accordingly. For more details on methods for setting institution-specific cut scores, see section 7.3.5. Though results for the ACT writing test are not shown in Table 7.8, a recent study involving 28 postsecondary institutions indicated that scores from the enhanced ACT writing test that was introduced in fall 2015 were related to the grades earned in English Composition, providing validity evidence in support of using the ACT writing test to help inform course placement decisions (Radunzel, 2019). Specifically, scores from the ACT writing test were found to contribute incrementally to the prediction of English Composition grades, beyond ACT English score and HSGPA.

Summary

The use of ACT scores for placement purposes increased the accuracy rate in all courses. The increases in accuracy rates were larger in math, social science, and natural science than they were in English courses. However, English courses tend to have higher percentages of students earning a B or higher, leading to smaller increases in accuracy rates. This phenomenon occurs regardless of the placement variable(s) used. Lastly, results varied across institutions for all the courses examined. Consequently, ACT encourages institutions to collect their own course outcome data and determine institution-specific placement cutoff scores, accordingly.

Table 7.8. Decision-Based Validity Statistics for Course Placement Using ACT Scores (Success Criterion = B or Higher Grade)

Course type		ACT score	No. of institutions	Median cut score*	Median logistic slope	Maximum accuracy rate			Increase in accuracy rate			Success rate		
						Q1	Med	Q3	Q1	Med	Q3	Q1	Med	Q3
English courses	Composition I	English	215	18	0.135	63	67	72	1	2	9	63	68	73
	Composition II		62	19	0.131	64	68	72	0	2	7	64	68	73
Math courses	Elementary algebra	Math	76	19	0.244	60	64	68	6	13	23	58	63	68
	Intermediate algebra		79	21	0.203	59	62	65	5	14	21	56	60	63
	College algebra		134	22	0.203	62	66	69	7	13	24	61	65	69
	Statistics/probability		17	21	0.184	61	65	68	3	10	22	59	61	68
	Precalculus		27	24	0.184	63	66	69	1	8	22	61	65	69
	Trigonometry		41	24	0.184	62	65	68	5	11	21	60	64	67
	Calculus		15	27	0.146	61	67	69	1	9	18	62	67	69
Social science courses	American history	Reading	60	23	0.114	61	63	66	4	11	19	59	61	64
	Other history		30	23	0.147	63	66	71	5	9	18	63	67	70
	Psychology		107	22	0.126	63	65	69	2	9	18	62	65	70
	Sociology		53	21	0.118	63	65	68	1	5	14	63	66	68
	Political science		33	22	0.108	61	62	65	3	6	15	62	64	66
	Economics		10	24	0.111	60	63	65	3	16	25	58	60	64
Natural science courses	Biology/life sciences	Science	108	23	0.196	63	65	69	6	14	24	60	64	67
	General chemistry		55	26	0.148	60	63	67	4	17	25	59	61	64

Note. Placement analyses that did not yield an optimal cutoff score (i.e., the logistic function did not include a probability of .50) were not summarized in this table.
*The median cut scores reported in the tables were weighted to reflect the national population of high school graduates to be consistent with the ACT College Readiness Benchmarks. Med = median.

7.3.3 Incremental Validity of ACT Scores and High School Grades in Course Placement

ACT encourages institutions to use multiple measures for placing students into college courses. Previous studies have reported that test scores and HSGPA, when used together, provide more information than either measure used alone (Noble, Schiel, & Sawyer, 2004; Sawyer, 2010). Specifically, the use of multiple measures often results in stronger predictive relationships with course grades and increased classification accuracy. Improved classification accuracy has important implications for institutions, especially at community colleges where large percentages of students enter college academically unprepared and require remediation (Sparks & Malkus, 2013). A recent study examined the joint use of ACT scores and HSGPA for course placement at community colleges to demonstrate how using multiple measures can result in more informed placement decisions (Westrick, 2016). This study concluded that institutions can more accurately predict a student's chance of success in college courses when they use more than one measure.

Supplemental analyses using the same data set were conducted to obtain the median accuracy rate (percentage of correct placement decisions), the median increase in accuracy rate (increment in percentage of correct placement decisions over using no predictors), and observed success rate (Table 7.9). For both English Composition I and College Algebra, the joint use of ACT test scores and HSGPA resulted in the highest accuracy rates, indicating that institutions can make better placement decisions if they use both ACT test scores and HSGPA together for placement. Additional information on the methodology used in these supplemental analyses can be found in another report by Westrick and Allen (2014) that conducted similar analyses using ACT Compass® scores instead of ACT scores before the ACT Compass test was retired.

Table 7.9. Median Placement Statistics for ACT Scores and HSGPA as Predictors at Community Colleges

Course	Predictor	Students	Median accuracy rate	Median increase in accuracy rate	Observed success rate
English Composition (259 Institutions)	ACT English	288,266	63.3	4.9	60.6
	HSGPA	256,110	66.7	8.3	61.2
	ACT English, HSGPA, & ACT English × HSGPA	256,100	66.8	7.9	61.2
College Algebra (182 Institutions)	ACT Math	132,850	66.5	25.9	42.2
	HSGPA	119,228	67.7	19.9	43.2
	ACT Math, HSGPA, & ACT Math × HSGPA	119,228	68.6	24.5	43.2

Note. Success rates varied across the three analyses for each course because the data sets were slightly different (not all students had both ACT scores and HSGPA data). Observed success rates (percentage of those with a B or higher grade) were calculated across all institutions combined. Accuracy rates were calculated at each institution.

7.3.4 Differential Prediction by Student Demographic Groups in Course Placement

A study by Allen (2016b) examined the predictive validity of using ACT scores for course placement by student demographic group. The study focused on four student demographic groups: English learners, students with disabilities, students of color (African American, Native American, Hispanic, Native Hawaiian, students of multiple races, and students of other races, not including White and Asian), and students from low-income households (<$36,000). Specifically, the study examined the extent that ACT cut scores associated with a 50% chance of earning a B or higher grade varied by demographic group.

The results of this study are consistent with prior research showing slight overprediction for students with disabilities (Huh & Huang, 2016; Ziomek & Andrews, 1996), students of color (Lorah & Ndum, 2013; Noble, Crouse, & Schultz, 1996; Sanchez, 2013; Sawyer 1985), and students from low-income households (Lorah & Ndum, 2013; Sanchez, 2013) and slight underprediction for English learners (Mattern, Patterson, Shaw, Kobrin, & Barbuti, 2008; Patterson & Mattern, 2012) when using standardized test scores to predict individual first-year course grades and overall FYGPA. Despite some of these differences, the accuracy rates at optimal ACT cutoff scores associated with predicting first-year course success were found by Noble et al. (1996) to be somewhat comparable across gender and racial/ethnic groups. Moreover, that research also identified similar patterns of overprediction and underprediction by gender and race/ethnicity when using high school subject-area GPAs alone to predict first-year college grades. Taken together, these findings highlight the importance of using multiple measures in making course placement decisions. This statement is further substantiated by a study showing that psychosocial constructs (motivation and self-regulation) helped to explain the gender gaps in first-year course outcomes that were observed after adjusting for ACT scores and the type and admission policies of the college the student attended (Ndum, Allen, Way, & Casillas, 2015).

7.3.5 Methods for Setting Cutoff Scores

Institutions have unique placement needs that require locally developed cutoff scores rather than the median optimal cutoff scores from the studies described above. This section provides guidance concerning procedures that institutions might follow to set cutoff scores. There are multiple ways to establish cutoff scores or decision zones for placement of students into different courses. The procedures for setting cutoff scores include the use of logistic regression and decision-based statistics, evaluation of local score distributions (often with respect to institutional resources), judgmental procedures based upon a content review of the items, and comparisons with reference populations.

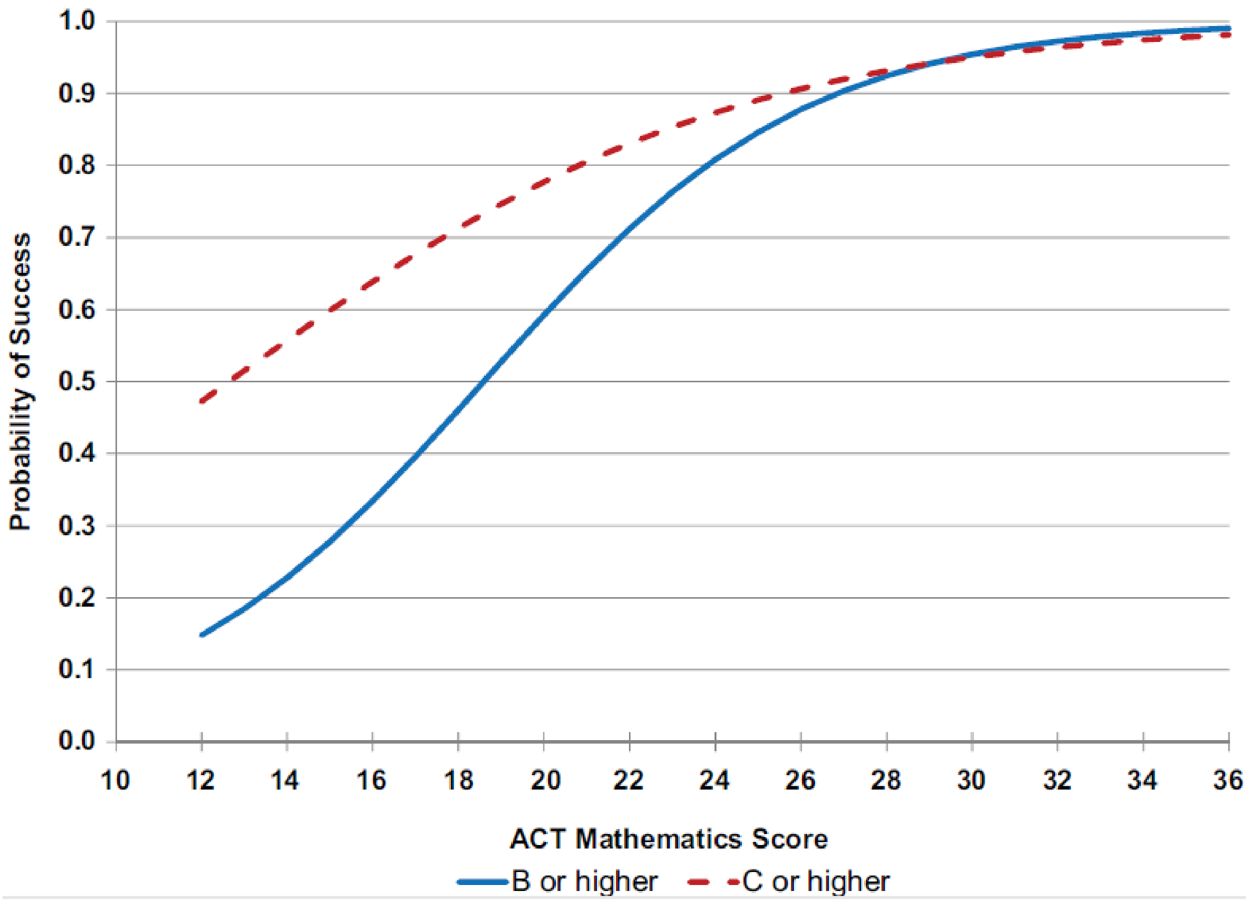
It is often advisable to interpret cutoff scores as guides rather than as rigid rules. One way to do this is to use decision zones. A decision zone is an interval around the cutoff score; students whose test scores (or other variable values) are in a decision zone are encouraged to provide more information about their academic qualifications and skill levels. For example, it might be appropriate to identify an ACT English score range of 17–20 as a placement decision zone for Composition courses. Students whose scores are above 20 would be placed into Composition. Those with scores below 17 would be placed into a developmental writing course that prepares them for Composition. Students whose scores fall into the decision zone would be advised that their skills appear to be on the borderline of readiness for Composition. Their option, with the advice of an advisor, would be to enroll in a developmental course (or participate in other appropriate skill-building services) to improve skills prerequisite for the Composition course or to enroll directly in the Composition course, with full awareness that most of the other students will probably have a stronger base of skills in the prerequisite areas. To provide more information about their readiness for Composition, another test of writing skills could be administered to the students whose scores fall into the decision zone.

A course placement study generates the probability of success, accuracy rate, success rate, and percentage not admitted or percentage placed in a lower-level course. If a test is effective for placement, then higher test scores should correspond to higher probabilities of success. Probability of success information can be used for advising individual students. It also serves as the basis for computing the group statistics used to validate tests and to select cutoff scores. As an example, Table 7.10 shows the relationship between students' ACT mathematics scores and their probability of earning a B or higher grade and a C or higher grade in Mathematics 100, a course at an institution. In this course, the probability of earning a grade of a B or higher corresponding to an ACT mathematics score of 18 is 0.46. That is, 46 out of 100 students with an ACT mathematics score of 18 would be expected to earn a B or higher grade in Mathematics 100. This information is also shown graphically in Figure 7.9.

Table 7.10. Probability of Success in Mathematics 100 Given ACT Mathematics Score

ACT math score	Probability of success (B or higher)	Probability of success (C or higher)
34	.98	.97
33	.98	.97
32	.97	.96
31	.96	.96
30	.95	.95
29	.94	.94
28	.92	.93
27	.90	.92
26	.88	.91
25	.85	.89
24	.81	.87
23	.76	.85
22	.71	.83
21	.65	.81
20	.59	.78
19	.53	.75
18	.46	.71
17	.40	.68
16	.33	.64
15	.28	.60
14	.23	.56
13	.18	.51
12	.15	.47

Figure 7.9. Probability of Success in Mathematics 100 Given ACT Mathematics Score



Decision-based statistics provide information about how a placement system affects groups of students. Such group-level information is important in validating and selecting cutoff scores for placement. The percentage of students who would be placed in lower-level courses is one important consideration. The availability of instructors, classrooms, and other resources affect how many students can be enrolled in either standard or lower-level courses. Moreover, if a test is effective for placement, then it should have a high estimated accuracy rate. That is, whether students are placed in a standard course or placed in a lower-level course, the decision should be correct more often than not. Finally, using an effective placement test should also result in a high estimated success rate, which means that most students placed into a course should be successful.

Table 7.11 is provided as an example of these statistics. If an ACT mathematics cutoff score of 20 were used for placement into Mathematics 100, then about 54% of the students would be placed into a lower-level course. With respect to the success criterion of a B or higher, about 69% of all the placement decisions (into either course) would be correct ones; of the students placed into Mathematics 100, about 76% of them would be expected to be successful.

The "optimal" cutoff score is a reasonable starting point for the selection process and can be found by identifying the score that corresponds to a probability of success of about 0.50. In

Tables 7.10 and 7.11, the ACT mathematics score of 19 is the cutoff score associated with at least a 50% chance of earning a grade of B or higher and the score that would maximize the accuracy of placement into Mathematics 100 (69%) for the B or higher success criterion.

One should keep in mind, however, that the cutoff score that maximizes the accuracy rate may be associated with a success rate and a percentage of students not admitted (or placed in the lower-level course) that is not acceptable to an institution. In Table 7.11, using the optimal cutoff (ACT mathematics score of 19) would place approximately 46% of the students into the lower-level course, and, with respect to the B or higher success criterion, about 73% of the students who would enroll in Mathematics 100 would be successful. A lack of resources may make it impossible for an institution to place 46% of their students into lower-level courses. A solution might be to use a cutoff score of 18. This would result in an accuracy rate nearly identical to the rate associated with a score of 19, but only 38% of the students would be placed into the lower-level course. The disadvantage of lowering the cutoff score would be that the percentage of students estimated to be successful in Mathematics 100 would decrease to 69%. The institution would need to consider the consequences of selecting alternative cutoff scores as they relate to resources, as well as to institutional goals and policies.

Table 7.11. Decision-Based Statistics for Placement Based on ACT Mathematics Score

ACT mathematics score	Percent placed in lower-level course	B or higher		C or higher	
		Estimated accuracy rate	Estimated success rate	Estimated accuracy rate	Estimated success rate
34	99	45	98	24	98
33	99	45	98	24	97
32	99	45	98	24	97
31	99	45	98	24	97
30	99	45	97	24	97
29	99	45	96	25	96
28	97	47	93	27	94
27	94	49	92	29	93
26	90	53	90	32	92
25	85	56	88	36	91
24	79	60	86	41	90
23	71	64	83	47	89
22	65	66	81	51	88
21	58	68	78	55	87
20	54	69	76	57	86
19	46	69	73	61	85
18	38	68	69	65	83
17	24	65	63	70	80
16	10	60	59	74	78
15	2	57	56	76	77
14	0	56	56	76	76
13	0	55	55	76	76

Local Score Distributions

Institutional personnel are often required to establish cutoff scores on the basis of administrative considerations (e.g., availability of instructional staff and facilities). Score distributions can be used under these conditions to provide preliminary cutoff scores. Cutoff scores based on score distributions are easy to communicate and to implement in placement systems. However, students' true abilities may be inconsistent with the selected cutoff score; that is, students who are underprepared for college may be incorrectly placed in the standard course. For more accurate decisions, ACT scores or other variables related to college or course outcomes should be used for placement decisions.

Expert Judgment

When expert judgment is used to establish cutoff scores, institutional personnel should conduct a thorough review of the test content. Based on this review, institutions may determine that a student correctly answering a certain percentage or more of the items has demonstrated sufficient knowledge of the subject to be placed in a particular course.

There are a variety of methods for determining the cutoff score associated with the minimum level of skills required for placement. (For a description of some of these methods, see Cizek & Bunch, 2006.) These methods often require content experts to judge how a “borderline” test taker (i.e., one whose knowledge and skills just barely reach the decision borderline) would perform on each item. Since each of these methods relies on subjective judgment, inspection of actual performance data is also recommended.

Other Comparison Populations

Cutoff scores can also be set using the scores from the ACT national norms or Table 7.8. This is particularly helpful when local normative data are not available. The normative distribution would be used in a manner similar to that described above for local score distributions. A student taking a specific test would be placed in a standard course if he or she scored at or above the scale score corresponding to a predetermined percentile rank in the score distribution of the reference population. Users should note that local distributions of ACT scores and grades may differ markedly from national distributions. Therefore, cutoff scores derived from national data should be validated and later adjusted as warranted when local data become available.

Monitoring Cutoff Scores

Once an institution selects a procedure and establishes a cutoff score, it is essential for the institution to continually monitor the effectiveness of the cutoff score. Experience may suggest adjusting established cutoff scores.

7.4 Using ACT Scores to Evaluate Students’ Likelihood of College Success in the First Year and Beyond

Sections 7.2 and 7.3 summarized the results of various studies that examined the relationships between ACT scores and first-year course grades for admission and placement decisions. This section describes studies illustrating the relationship between college readiness as measured by the ACT and students’ success using additional outcomes from the first year of college and beyond. The information presented in this section may be considered further evidence supporting the use of ACT scores for making college admissions decisions. The first subsection focuses on relating ACT Benchmark attainment to first-year outcomes that include college enrollment, first-year college grades, and college retention. The second subsection focuses on relating ACT scores to ACT Collegiate Assessment of Academic Proficiency (CAAP) scores taken by students during their second year of college. The third and fourth subsections focus on relating ACT scores to longer-term outcomes that include cumulative college GPA at graduation and degree attainment. The fifth subsection focuses on relating the ACT STEM score to students’ chances of persisting and completing a college degree in a STEM-related field.

7.4.1 Statistical Relationships between College Readiness and First-Year College Success

This section provides estimates of students’ chances of college success for several different first-year outcomes examined by ACT College Readiness Benchmark attainment in individual subject areas as well as by the number of Benchmarks met (see Chapter 5 or Allen, 2013, for a description of the Benchmarks). Using more recent freshman cohorts, the results presented here update some findings from an earlier study conducted by ACT (2010).

Data and Method

College outcomes included enrollment into any college the fall following high school graduation, earning a B or higher grade in first-year college courses, achieving a FYGPA of 3.0 or higher, and remaining enrolled at the initial institution in year two. College readiness was measured by ACT College Readiness Benchmark attainment.

College enrollment rates were based on approximately 1.9 million high school students who took the ACT and indicated that they would graduate from high school in 2015. Colleges included both two-year and four-year institutions. College retention rates were based on approximately 1.3 million ACT-tested students from the 2015 graduating class who enrolled in a postsecondary institution the fall following high school graduation, according to the National Student Clearinghouse database. More than 2,800 colleges were included. Data for FYGPA included approximately 430,000 ACT-tested students from nearly 300 postsecondary institutions who participated in research services offered by ACT. Approximately 125,000 students were included in the analysis for English Composition I; 31,000 for English Composition II; 20,000 for Intermediate Algebra; 69,000 for College Algebra; 5,000 for Precalculus/Finite Math; 18,000 for Calculus; 41,000 for American History; 77,000 for Psychology; 32,000 for Biology; and 31,000 for Chemistry. For all outcomes except college enrollment, hierarchical logistic regression models were used to estimate students' chances of success as a function of Benchmark attainment or the number of Benchmarks met, while statistically controlling for the institution attended. Random intercept models were estimated. For college enrollment, observed rates were calculated.

Results

Students who met the ACT College Readiness Benchmarks were more likely than those who did not to: (a) enroll in college the fall following high school graduation (Figure 7.10; by 23 to 29 percentage points); (b) earn a B or higher grade in first-year college courses (Figure 7.11; by 18 to 27 percentage points); (c) achieve a FYGPA of 3.0 or higher (Figure 7.12; by 23 to 27 percentage points); and (d) remain enrolled at the same institution in year two (Figure 7.13; by 6 to 9 percentage points). Moreover, as the number of Benchmarks attained increased, students' likelihood of success also increased for each of the first-year outcomes examined (Table 7.12). For example, students' chances of enrolling in college increased from 45% for those who met none of the Benchmarks to 83% for those who met all four Benchmarks.

Figure 7.10. College Enrollment Rates by ACT College Readiness Benchmark Attainment

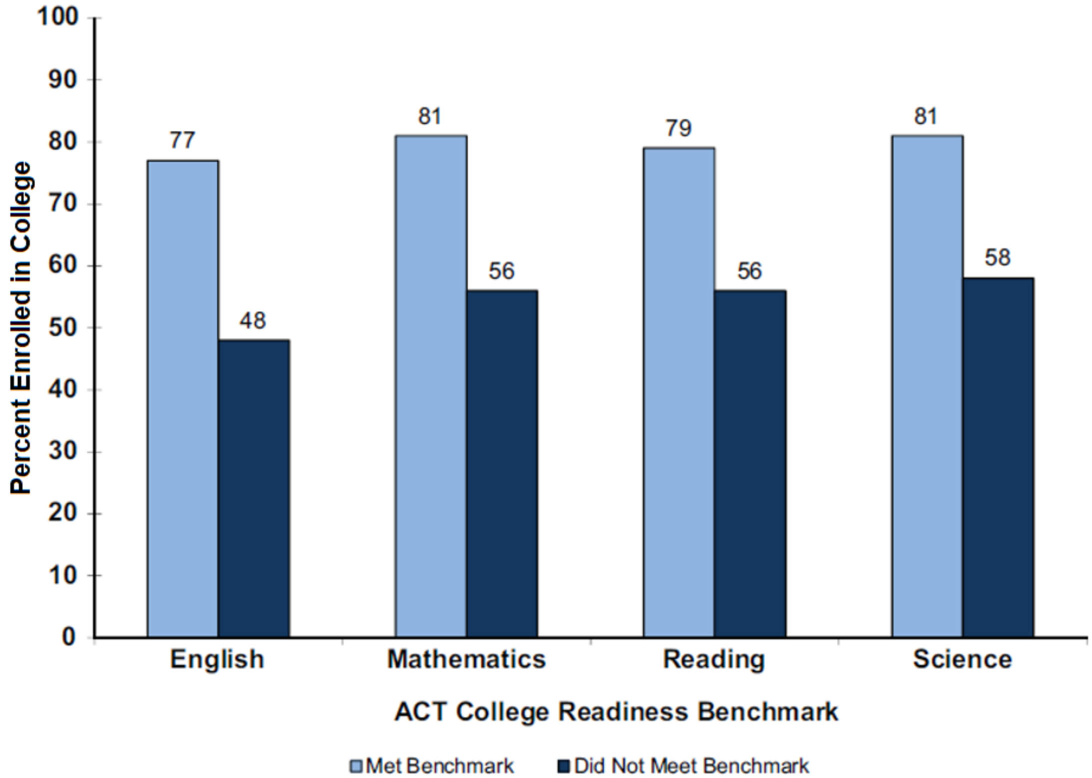


Figure 7.11. Students' Chances of Earning a B or Higher Grade in First-Year College Courses by ACT College Readiness Benchmark Attainment at a Typical Institution

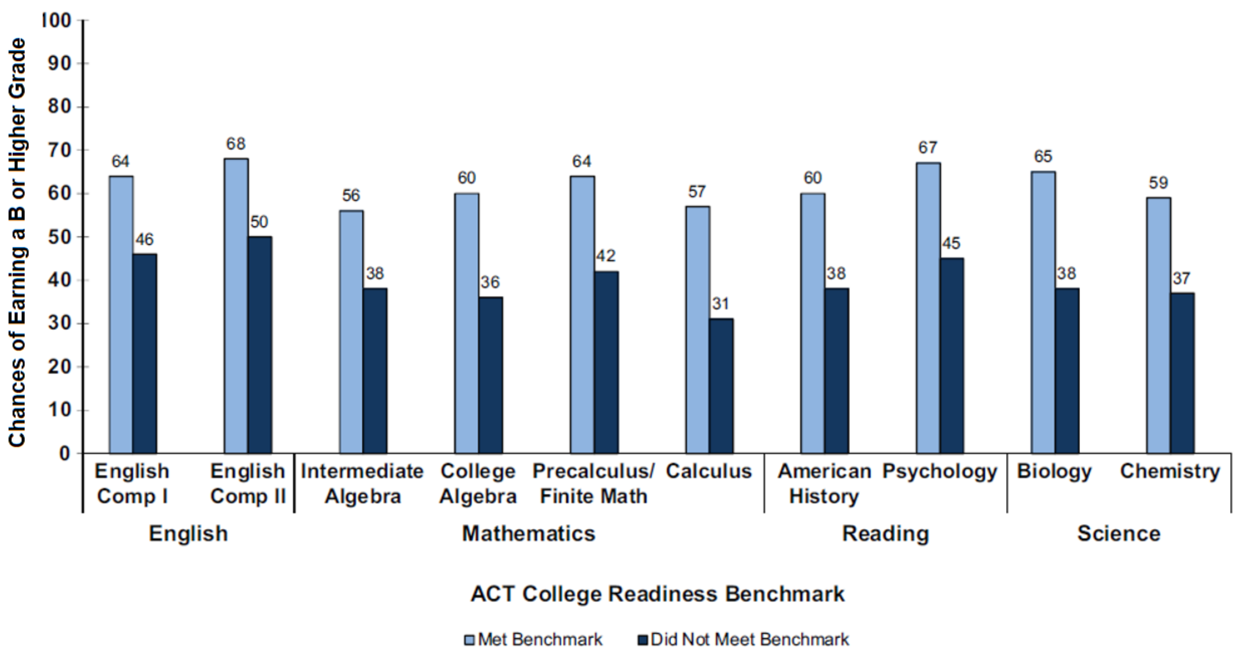


Figure 7.12. Students' Chances of Achieving a 3.0 or Higher FYGPA by ACT College Readiness Benchmark Attainment at a Typical Institution

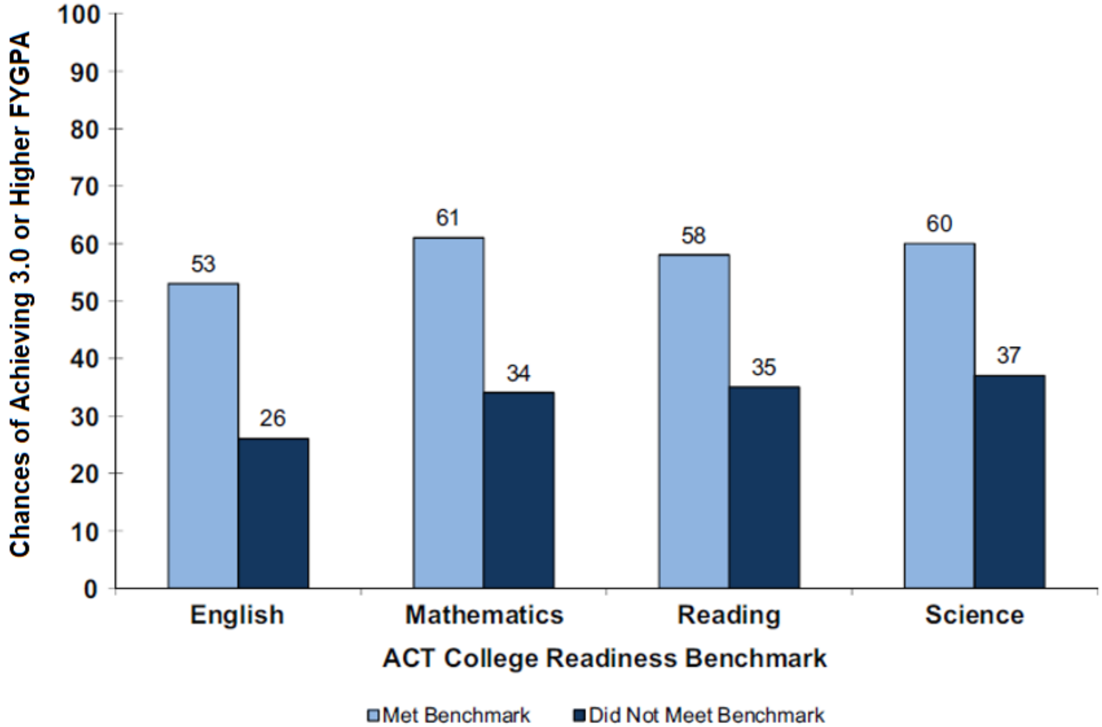


Figure 7.13. Students' Chances of Remaining Enrolled at the Initial Institution in Year Two by ACT College Readiness Benchmark Attainment

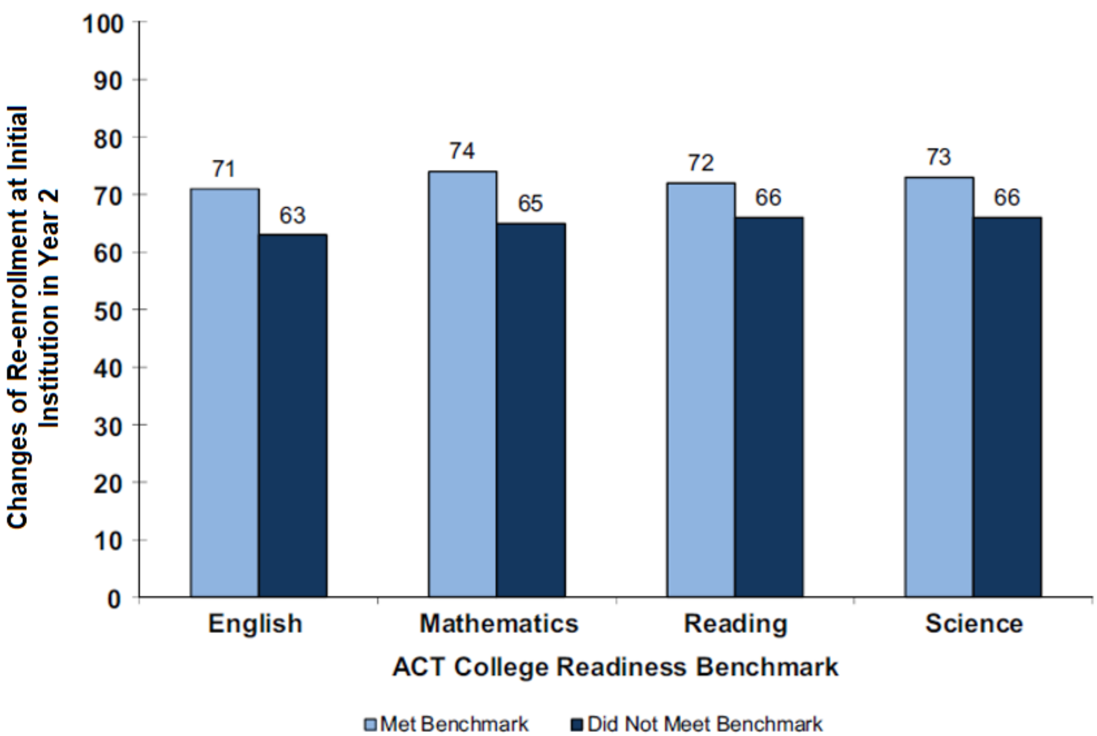


Table 7.12. Percentage Chance First-Year College Outcomes by Number of ACT College Readiness Benchmarks Met

Outcome	Number of ACT Benchmarks met				
	0	1	2	3	4
Enrollment	45	66	73	78	83
B or higher grade in course in					
English Composition I	45	54	61	68	75
English Composition II	51	58	65	71	76
Intermediate algebra	33	39	46	52	58
College algebra	29	37	46	55	64
Precalculus/finite math	38	45	52	60	66
Calculus	25	33	42	51	61
American history	29	40	51	62	72
Psychology	35	47	58	69	79
Biology	22	33	47	62	75
Chemistry	21	30	41	53	65
FYGPA of 3.0 or higher	22	33	45	57	69
Retention	62	66	69	73	76

Summary

The ACT College Readiness Benchmarks are good indicators of whether students have acquired the knowledge and skills to be successful in first-year college courses. The results from the current analyses also show that students who are better prepared academically for college (as indicated by meeting the Benchmarks) are more likely than less prepared students to immediately enroll in college and, once they enroll, tend to be more successful during their first year of college and to remain enrolled at their initial institution in year two.

7.4.2 Statistical Relationships Between ACT and ACT CAAP Scores

The previous section showed that students who are better prepared academically, as measured by meeting the Benchmarks, are more likely to succeed during their first year of college than less prepared students. In this section, to better understand the relationship between college readiness and student academic success into the second year of college, the relationships between ACT scores, Benchmark attainment, and ACT Collegiate Assessment of Academic Proficiency (CAAP) scores were examined for second-year college students. ACT CAAP was a standardized assessment program used by colleges and universities to evaluate student learning outcomes in their general education programs. ACT CAAP offered six independent test modules: reading, science, critical thinking, mathematics, writing skills, and writing essay (ACT, 2015a). The ACT CAAP assessment was taken by students between the academic years 2008–2009 and 2014–2015.

Data and Method

The sample included more than 16,000 college students who took ACT CAAP during the spring term of their second year and the ACT test in high school during their junior or senior year.

Because of the modular nature of ACT CAAP, not all students had all ACT CAAP scores. The results for English/writing skills were based on 11,221 ACT/CAAP-tested students. Results for the other subject areas were based on 11,892 students for mathematics, 10,574 students for reading, and 9,005 students for science. Self-reported cumulative college GPAs at the time of ACT CAAP testing were also available as an indicator of college achievement.

College readiness was measured by ACT College Readiness Benchmark attainment (see Chapter 5 or Allen, 2013, for a description of the Benchmarks). Descriptive statistics including means, standard deviations, percentages, and correlations were used to examine how ACT scores or Benchmark attainment related to ACT CAAP scores and cumulative college GPA in the second year.

Results

ACT scores were strongly correlated with ACT CAAP performance (Table 7.13). In addition, students meeting the ACT College Readiness Benchmarks in high school had higher average ACT CAAP scores than students not meeting the Benchmarks (Table 7.14). This pattern was observed in all four content areas. The difference in average ACT CAAP scores was as much as 6.6 points (on a 40 to 80 score scale). Moreover, as shown in Figure 7.14, students who met the Benchmarks in high school were more likely to have a cumulative college GPA greater than 3.0 in their second year of college.

Summary

These findings suggest that the use of ACT College Readiness Benchmarks can assist in determining who will succeed in college, even into the second year.

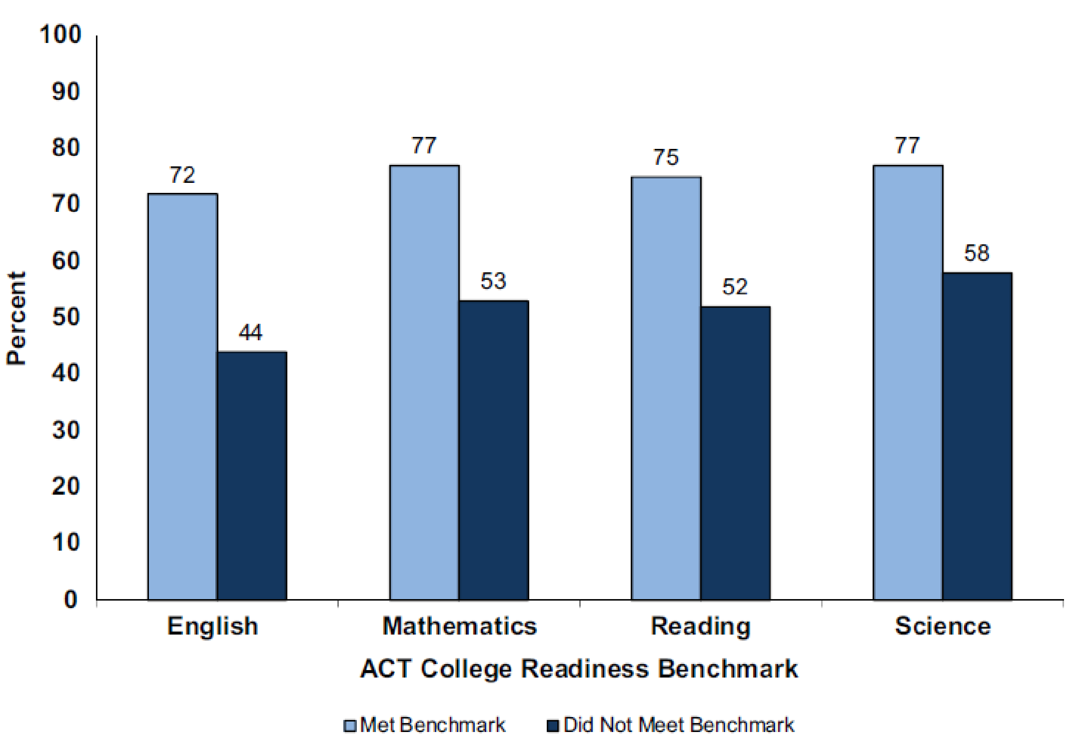
Table 7.13. ACT/CAAP Test Score Correlations

English/ Writing skills	Math/Math	Reading/Reading	Science/Science
0.77	0.73	0.70	0.67

Table 7.14. Average ACT CAAP Test Score by ACT Benchmark Attainment

ACT/CAAP content area		ACT Benchmark attainment	
		Met	Not met
English/Writing skills	Mean (SD)	65.0 (4.0)	58.4 (3.5)
	Number of students	8,418	2,803
Math/Math	Mean (SD)	60.7 (3.5)	55.9 (3.0)
	Number of students	5,145	6,747
Reading/Reading	Mean (SD)	64.3 (4.6)	58.1 (4.2)
	Number of students	5,199	5,375
Science/Science	Mean (SD)	63.8 (3.9)	58.5 (3.8)
	Number of students	3,514	5,491

Figure 7.14. Percentages Earning a Cumulative College GPA Greater Than 3.0 by ACT College Readiness Benchmark Attainment for Students Taking ACT CAAP During Sophomore Year and the ACT in High School



7.4.3 Statistical Relationships between ACT Scores and Cumulative College GPAs

A study by Tracey and Robbins (2006) examined the relationships between performance on the ACT and cumulative college GPA across time. Hierarchical linear modeling (HLM) was used to examine the relationship between ACT scores and college GPA while accounting for the nesting of students within colleges. The findings from this study indicated that performance on the ACT was predictive of cumulative college GPA across time. The researchers also examined how congruence measures between students' interests (as measured by the ACT Interest Inventory) and college major choice relate to college performance. For more details, see the full research article (Tracey & Robbins, 2006).

7.4.4 Statistical Relationships between ACT Scores and Degree Completion

Long-term student success is an important goal for students and postsecondary institutions. A study by Radunzel and Noble (2012) examined the relationships between performance on the ACT and degree completion at both two- and four-year institutions. Such information might be useful for early identification of students who could possibly benefit from additional academic and student support services upon entering college.

Data and Method

Data for this study included approximately 194,000 ACT-tested students who enrolled in college as first-time entering students in fall 2000 through 2006. Approximately 126,000 students who began at one of 61 four-year institutions were tracked for at least six years, and nearly 68,000 students who began at one of 43 two-year institutions were tracked for at least three years. The outcomes were bachelor's degree completion within six years from the initial institution for students beginning at four-year institutions and associate's degree completion within three years from the initial institution for students beginning at two-year institutions. Because many students beginning at a two-year institution transfer to a four-year institution without earning an associate degree (Radunzel, 2012), associate degree completion or transfer to an in-state four-year institution within three years was also evaluated for students beginning at two-year institutions. The latter outcome was evaluated for a subset of the two-year data from two state systems where students could be tracked across both two- and four-year institutions. Hierarchical logistic regression models were used to estimate institution-specific probabilities of degree completion based on ACT scores alone and in combination with self-reported HSGPAs. The accuracy rates and increases in accuracy rates over not using the predictor were calculated at the predictor value(s) associated with a 50% chance of degree completion (for more details on these decision-based statistics, see Section 7.2.1).

Results

As shown in Figure 7.15, as ACT Composite score increased, students' chances of completing a degree increased for both two- and four-year students. Additionally, as ACT Composite score increased, two-year students' chances of completing an associate degree or transferring to a four-year institution increased. As an example of the increase for those beginning at a four-year institution, students' chances of completing a bachelor's degree in six years was 41% for those with an ACT Composite score of 20, and it was 67% for those with an ACT Composite score of 30. Higher values of HSGPA were also associated with increased chances of degree completion.

The typical maximum accuracy rate and increase in accuracy rate across institutions associated with using ACT Composite score to predict bachelor's degree completion within six years were 64% and 24%, respectively. Similar rates were associated with using HSGPA alone (65% and 23%). In comparison, the typical maximum accuracy rate associated with using both predictors jointly was 2 to 3 percentage points higher than those based on the single-predictor models.

Figure 7.16 provides the estimated probabilities of completing a bachelor's degree within six years associated with different values of HSGPA and ACT Composite score. The figure illustrates the incremental usefulness of ACT scores beyond HSGPA for predicting who is likely to complete a degree. As both HSGPA and ACT Composite score increased, probabilities of success also increased. The ACT Composite score differential was larger for students with higher HSGPAs than those with lower HSGPAs. The same was true for the HSGPA differential when comparing students with higher and lower ACT Composite scores.

Summary

Both ACT Composite score and HSGPA were effective for predicting long-term college success at two- and four-year institutions. Other outcomes examined in the study included progress to degree completion (based on cumulative hours earned) and cumulative GPA at degree completion. Across the outcomes, ACT test scores increased prediction accuracy over that for HSGPA alone. The study also indicated that ACT Composite scores and HSGPA were primarily indirectly related to subsequent college outcomes through FYGPA. For additional information on this study, see the full report (Radunzel & Noble, 2012).

Figure 7.15. Probability of Degree Completion Based on ACT Composite Score (Radunzel & Noble, 2012)

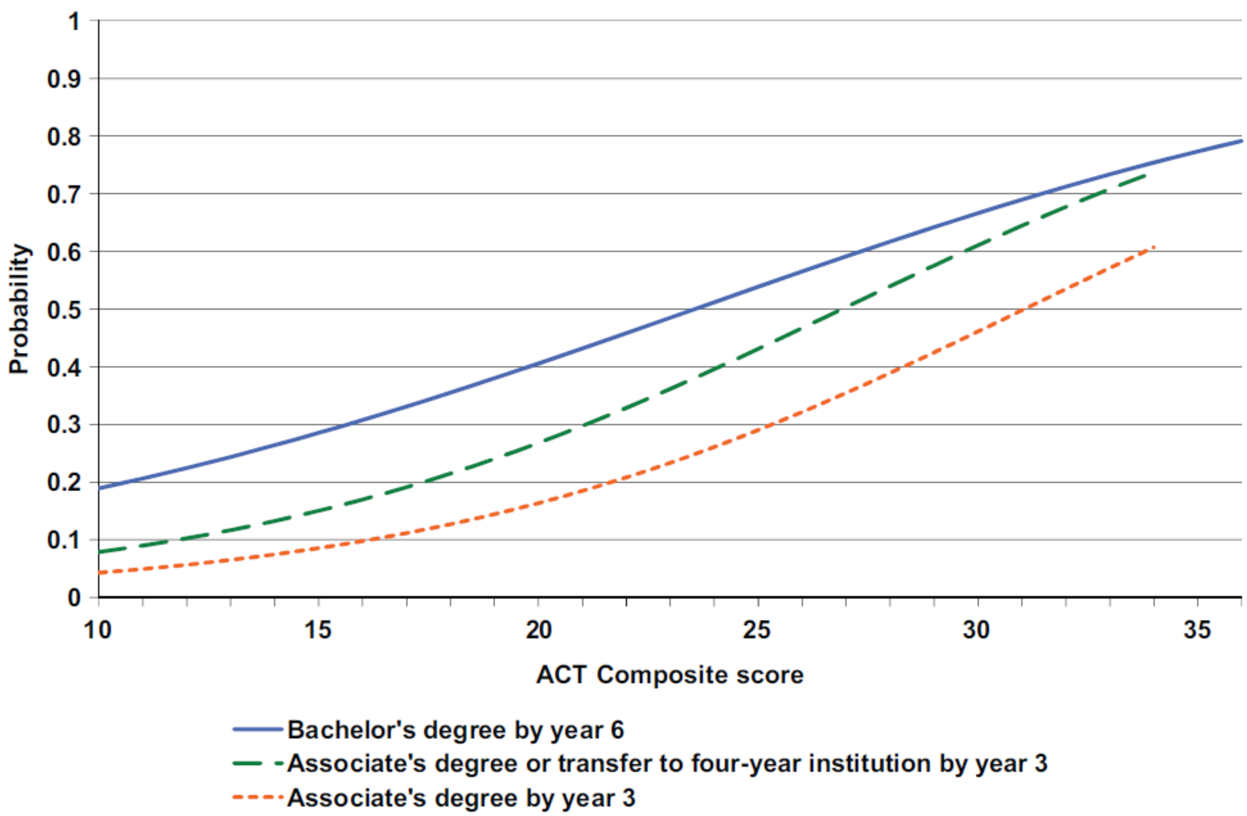
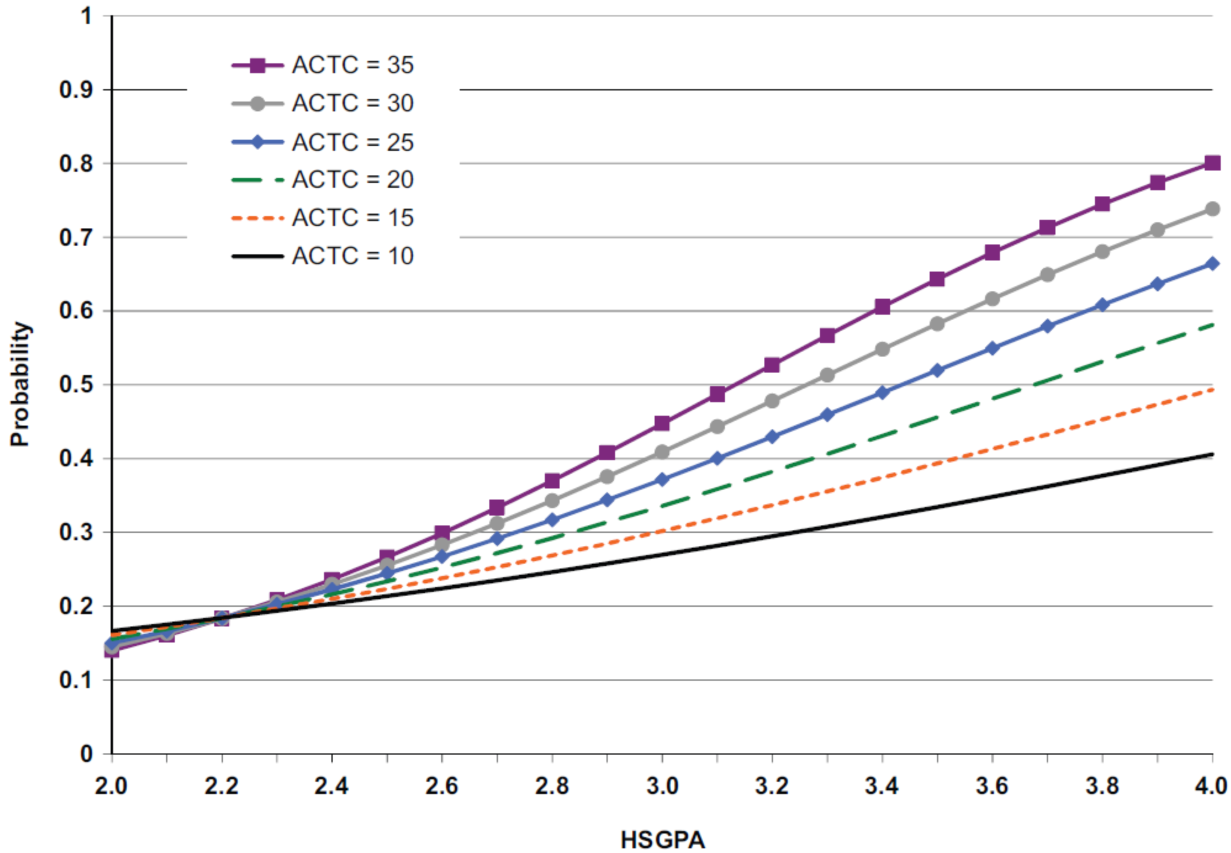


Figure 7.16. Probability of Bachelor's Degree Completion Within 6 Years by HSGPA and ACT Composite Score (Radunzel & Noble, 2012)



7.4.5 Statistical Relationships between ACT STEM Scores and Students' Chances of Succeeding in a STEM-Related Major

A study by Radunzel, Mattern, Crouse, and Westrick (2015) examined the ACT STEM score in relation to the likelihood of succeeding in a variety of STEM-related college outcomes: cumulative GPA over time, persistence in a STEM major, and ultimately completing a STEM degree. The results from the study illustrated that predicting student success in STEM-related fields is a valid use of the ACT STEM score.

Another study by Radunzel, Mattern, and Westrick (2016) suggested that the positive relationship between ACT STEM scores and students' chances of succeeding in a STEM major held even after statistically controlling for other student characteristics, such as high school coursework taken and grades earned, vocational interests, and demographic characteristics. This finding was consistent with a growing body of literature indicating that educational success is a not just a product of academic skills and knowledge but also of noncognitive factors such as motivation, academic goals, and academic self-efficacy (Mattern, Burrus, Camara, O'Connor, Hanson, Gambrell, Casillas, & Bobek, 2014).

7.5 Using ACT Scores to Assist with Program Evaluation

The ACT tests were developed to measure academic skills and knowledge that are learned in high school and are necessary for academic success in the first year of college. Validity evidence for using the ACT as a measure of educational achievement is documented at the beginning of this chapter. Since the ACT measures important educational outcomes, it might be considered for use in evaluating the effectiveness of school programs.

Before using the ACT in program evaluation, a school should conduct a content review to determine the extent to which the tests represent important outcomes the school wishes to measure. If there is a content match between the ACT and important local educational outcomes, the ACT may be considered as one component of a program evaluation system. ACT scores should not be relied on exclusively as evidence of program effectiveness. Rather, ACT scores should be considered with other indicators of program effectiveness routinely collected by schools.

Several cautions must be kept in mind when using the ACT for program evaluation. ACT-tested students may not represent all students enrolled in the school, so ACT results may not generalize to other populations of students at the school. That is, expectations of and conclusions drawn about a select group of students who take the ACT may differ from those concerning a larger group of college-bound students or those of the high school graduating class as a whole. In cases where the school administers the ACT to all their juniors or seniors through statewide or districtwide testing programs, this is less of a concern. Additionally, a school's average ACT scores can fluctuate from year to year, as evidenced in a study by Sawyer (2013b). In that report, Sawyer described simple ways for school officials to better understand whether yearly changes or trends over time in average ACT scores for their school are unambiguous instead of plausibly due to chance. Finally, without some measure of student achievement earlier in high school, judgments about educational development and achievement during high school may be misleading. This issue can be addressed by using the ACT in conjunction with PreACT, ACT's tenth-grade assessment program (ACT, 2020d), or PreACT 8/9 (ACT, 2020c).

7.5.1 Using ACT Scores as Outcomes for Program Evaluation

ACT scores can be used in various ways for program evaluation. A school could establish expected levels of educational achievement on ACT scores or Benchmark attainment for individual students, for the entire group of tested students, or for groups of students defined by common academic interests, high school coursework, or some other characteristic. Expected and actual levels of educational achievement could then be compared to evaluate program effectiveness.

In establishing expected levels of achievement for groups of students, several factors need to be considered, including the availability of resources both within and external to the school, the social climate of the school, the characteristics of the students from the school who take the ACT, and the level of students' academic preparedness upon entering the school.

One way to determine expected levels of educational achievement is by estimating them with growth models that use prior measures of achievement from earlier grades. For more details on using growth models with the ACT to evaluate program effectiveness, see Chapter 8.

7.5.2 Using ACT Scores as Measures of Prior Achievement for Program Evaluation

ACT scores may be used as measures of prior achievement to statistically control for differences among program participation groups when evaluating the effectiveness of specific educational programs in observational studies. Two examples are provided in this section.

The first example is a study that Noble and Sawyer (2013) conducted to examine whether taking developmental courses in college benefits students, in the sense that they are more successful than they would have been if they had not taken developmental courses. Students' chances of success for a variety of college outcomes (including grade in the standard-level course) were estimated using ACT test score, enrollment status (full- or part-time), and college type (two- or four-year) for students who first took the developmental course followed by the standard-level course and those who directly enrolled into the standard-level course. Students' chances of success were then compared between the two groups over the range of overlapping ACT scores. Like other studies of this kind, results indicated that students who took developmental courses were less successful as a group than those who did not take developmental courses with respect to GPA and persistence over time, as well as degree completion within a fixed time period.

The second example comes from two studies undertaken to compare the short- and long-term college outcomes between students who had taken dual-credit/dual-enrollment coursework in high school and those who had not (Crouse & Allen, 2014; Radunzel, Noble, & Wheeler, 2014). The studies revealed that students entering with dual-credit hours were generally more academically able (as measured by ACT scores and high school rank) than students who had not taken dual credit in high school. Students in the two groups also differed on other student and school characteristics related to college success. After statistically controlling for ACT scores and other student and school characteristics, the findings suggested that students entering college with dual-credit/dual-enrollment credit performed as well as those with no dual credit in terms of the college grades earned in subsequent courses taken in college. This was despite concerns that dual-credit courses may not sufficiently prepare students for subsequent college courses. Additionally, dual-credit students were generally found to be more likely to complete a college degree in a timely manner. These studies demonstrate the value of including ACT scores as measures of prior achievement in program evaluation studies. For more details on the examples provided, see the full reports.

Chapter 8

Growth Models Using ACT Test Scores

Understanding student growth models can help students, parents, educators, and practitioners make better use of ACT® data. Growth models can be used to answer several important questions:

- How does the growth of students from my school compare to national growth averages?
- How much do my students need to grow to reach their ACT score goals?
- How much do ACT test scores typically increase over a one-year period?
- Which high school courses have the strongest relationships with student growth?

Growth models that incorporate scores from various ACT assessments (e.g., the PreACT and the ACT test) can be used to measure progress—both for individuals and groups of students. Measures of student growth can be used to inform teaching practices and to assess the effectiveness of new programs and interventions. In this chapter, gain-based models will first be distinguished from conditional status models. Subsequent sections will discuss resources that are available for implementing growth models based on the ACT test, summarize research explaining variation in student growth, discuss using growth models for evaluation of programs and school effectiveness, and summarize research on ACT test-retest statistics. The interested reader should also examine *ACT Growth Modeling Resources* provided online, which include lookup tables for expected changes in test scores for different ACT assessments and grade levels.

8.1 Distinguishing Gain-Based Models from Conditional Status Models

There are several different methods for describing student- and group-level growth—including methods based on gain scores, trajectories, achievement level transitions, residual gains, projections, conditional growth percentiles, and multivariate models (for a description of each type of growth model, see Castellano and Ho, 2013). These methods are classified by their underlying statistical foundations into one of three categories: gain-based models, conditional status models, and multivariate models (Castellano & Ho, 2013). ACT test scores can be used within all three categories of growth models. However, as described in this chapter, the ACT most directly supports gain-based and conditional status models.

Gain-based and conditional status models support contrasting perspectives on growth. Understanding the two models is essential for accurate selection and use of growth models. Gain-based and conditional status models are fundamentally different in two ways: statistical foundation and reliance on common score scales.

8.1.1 Statistical foundation

A gain score is the arithmetic difference between two scores at different time points. Gain-based models express growth as the difference in test scores over time and are meant to answer the question: "How much has a student learned on an absolute scale?" (Castellano & Ho, 2013, p. 35). Gain scores can be extrapolated to future time points to support growth predictions. Trajectory models are a type of gain-based model meant to answer the question: "If this student continues on this trajectory, what will they likely score in the future?"

In contrast, conditional status models address the question: "How well did a student score, relative to peers with similar score histories?" While gain-based models attempt to describe growth in an absolute sense, conditional status models attempt to describe growth relative to peers. Conditional status models support normative interpretations of student growth.

Conditional status models often use regression methods that establish expectations for student test scores, based on their past scores. Comparing actual test scores to expected test scores allows users to determine if students have met expectations for "normal" growth. Popular forms of the conditional status model include the student growth percentile (SGP) model and the residual gain model. Similar to gain-based models, conditional status models can be used to describe student growth and to predict future test scores. The SGP model was implemented for the *ACT Growth Modeling Resources*, as discussed later in this chapter.

8.1.2 Reliance on Common Score Scales

Gain-based models require that test scores from multiple time points share a common scale. This can be achieved through vertical scaling (e.g., test scores from two grade levels placed on the same scale) or by using the same test at multiple time points. When the tests share a common scale, the difference in test scores is meaningful, enabling gain-based models. The PreACT® and ACT tests are examples of tests that are vertically scaled.

In contrast, conditional status models do not require that the tests have a common scale. For example, ACT Aspire and the ACT test do not share a common scale, but conditional status models can still be used to describe growth for students who took ACT Aspire® and the ACT test. Conditional status models are often operationalized using regression methods. Expectations for test scores (Y) are based on a prior test score (X) or set of prior test scores. As in regression, there is no requirement that Y and X be on the same scale. Test scores that are on a common scale can be used within both gain-based models and conditional status models. Conditional status models are flexible in that the model can use prior year scores from a single year or a collection of scores from multiple prior years.

8.2 ACT Growth Modeling Resources

To help users implement growth models based on different ACT assessments, ACT provides SGP lookup tables for various assessment combinations and grade levels. ACT's Growth Modeling Resources (https://www.act.org/content/act/en/research/services-and-resources/act-growth-modeling-resources.html) are updated periodically to reflect changes in student growth over time and to add more combinations of assessments and grade levels.

The SGP model describes a student's current achievement compared to other students with similar prior achievement scores. The SGP model expresses growth as a percentile rank relative to "academic peers." The SGP is meant to answer the question: "What is the percentile rank of a student's current score compared to students with similar score histories?" For example, a student earning an SGP of 75 performed as well as or better than 75 percent of her or his academic peers with similar score histories. SGPs provided by ACT are expressed as whole number values from 1 to 100.

Like other conditional status models, the SGP model accommodates multiple prior test scores (in the same subject or from different subjects) and does not require test scores from multiple time points to share a common scale. SGPs are often calculated using quantile regression (Koenker, 2005). This method for calculating SGPs does not require linear relationships between prior and current test scores, nor does it require constant variance across prior scores. Software that estimates SGPs using quantile regression methods is open-source and is available in the SGP package for R (Betebenner, VanIwaarden, Domingue, & Shang, 2014).

Many states and school systems use the SGP model to describe student growth, predict future test scores, and examine differences in growth across student groups (e.g., by race/ethnicity, gender, economic status, and disability status). Measures of aggregate growth include the mean and median SGP. Research suggests that mean SGP may have advantages over the median SGP in terms of efficiency, greater alignment with expected values, and greater robustness to scale transformations (Castellano & Ho, 2015).

The mean SGP can be used to identify relative growth differences across classrooms, schools, districts, and other groups of interest. When comparing mean SGPs across groups, it is important to consider whether differences in the composition of the groups could explain differences in mean SGP. For example, a school serving students from low-income families might be expected to have lower mean SGP than a school serving students from higher-income families.

The *ACT Growth Modeling Resources* include SGP lookup tables that can be used to find the SGP value (ranging from 1 to 100) associated with each combination of current-year test score and prior-year test score. The lookup tables provide an estimate of the SGP for each possible combination of same-subject test scores for various growth periods. When interpreting SGPs, the reference group used to estimate the model should always be considered. As of fall 2021, SGP lookup tables available for the ACT test include:

- ACT Aspire to ACT Aspire (one-year growth): The sample group consists of examinees who tested in consecutive years between grades 3 and 10.
- ACT Aspire to ACT Aspire (two-year growth): The sample group consists of examinees who tested two years apart between grades 3 and 10.

- ACT Aspire to PreACT (one-year growth): The sample group consists of examinees who took ACT Aspire in grade 9 and the PreACT in grade 10.
- ACT Aspire to ACT (one-year growth): The sample group consists of examinees who took ACT Aspire in grade 10 and the ACT in grade 11.
- PreACT to ACT (one-year growth): The sample group consists of examinees who took the PreACT in grade 10 and the ACT in grade 11.
- ACT to ACT (six-month growth): The sample group consists of examinees who took the ACT test in grades 11 and 12.

SGPs are provided for English, mathematics, reading, science, ELA, STEM, and Composite.

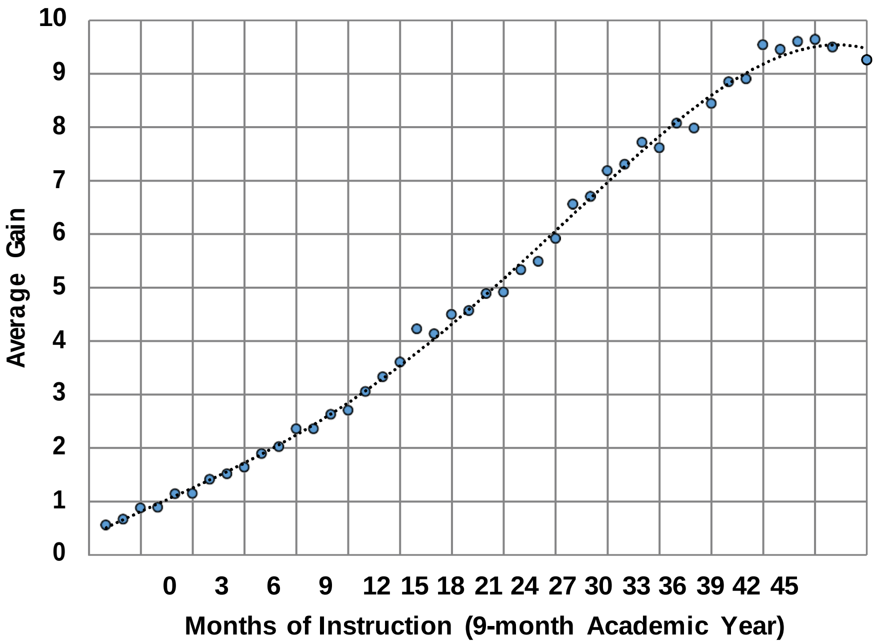
8.3 Explaining Variation in Student Growth

Academic growth based on ACT test scores varies across students, schools, and other student groups. Some of the variation in growth can be explained by factors such as instructional time and high school coursework and grades. This section summarizes research that explains some of the variation in student growth.

8.3.1 ACT Score Gains by Months of Instruction

Camara and Allen (2017) examined the relationship between instructional time and changes in ACT scores using longitudinal data. The sample included over 2.8 million test-retest instances for students from the 2016 ACT-tested graduating class. This research captured typical test-retest periods (e.g., April grade 11 to October grade 12) and much longer test-retest periods (e.g., grade 7 to grade 12), enabling an examination of ACT score gains across multiple years of instruction. Results indicated that ACT scores steadily increased with more instructional time (Figure 8.1). ACT Composite scores generally increased by 0.20 to 0.25 points per month of instruction, though the increment was larger for shorter periods (1–3 months), perhaps due to practice effects. Over a 4-year period (36 months of instruction), students gained about 8.5 ACT Composite score points, on average.

Figure 8.1. Average Gain in ACT Composite Score by Months of Instruction



Note. Months of instruction are based on time between ACT tests, not counting summer months (June, July, and August).

8.3.2 Predictors of Long-Term Growth

Estimated Benefits of Additional High School Coursework and Improved Course Performance in Preparing Students for College

Strategies for increasing academic growth and improving college readiness include taking more rigorous college-preparatory courses and expending more effort in these courses. Sawyer (2008) examined how taking additional courses and earning higher grades were associated with high school students' academic preparation for college, using data from students who took ACT Explore in 8th grade, ACT PLAN® in 10th grade, and the ACT in 11th or 12th grade.

Students' background characteristics, ACT Explore scores, high school attended, high school coursework, and high school grades were all related to ACT scores, but ACT Explore scores were by far the most strongly related. Taking more standard or advanced courses in high school and earning higher grades were associated with higher ACT scores even for students who had high ACT Explore scores. There was significant variation in high schools' average ACT scores, even after accounting for differences in their students' characteristics. The apparent benefit of additional standard coursework, advanced/honors coursework, and higher grades also varied significantly among high schools.

Bassiri (2014) replicated Sawyer's study using a more recent cohort of students (high school graduates of 2013 who took ACT Explore in 8th grade) and using updated values for the College Readiness Benchmarks. The source data contained records for 399,642 students from 6,228 high schools. In contrast to Sawyer's (2008) study, accelerated, honors, or advanced courses were excluded from the predictive models due to having large percentages of missing data (39%–56%). Furthermore, dummy variables corresponding to statewide testing for eight more states in addition to Colorado and Illinois were included in the models as potential Level-2 predictors of the intercept. In general, the findings were consistent with the earlier study (Sawyer, 2008). The few exceptions included coursework in English and coursework in social studies being significant predictors of ACT English and ACT reading scores, respectively.

Predictors of Academic Growth in Secondary School among Academically Advanced Youth

Many academically advanced youth take the ACT test in 7th grade for academic talent searches and again in 11th or 12th grades for college admissions (Allen, 2016a), enabling an investigation of predictors of growth during secondary school. Wai and Allen (2019) tested whether variation in academic growth among academically advanced youth is explained by socio-demographics, high school characteristics, coursework taken, high school GPA, Holland-type vocational interests (Holland, 1997), and extracurricular activities.

The study examined predictors of student growth over a long period: 7th grade to 11th or 12th grade. Predictors of growth included malleable factors such as high school coursework and grades as well as background variables such as race/ethnicity and family income. All variables combined—socio-demographics, interests, high school characteristics, high school coursework and GPA, and extracurricular activities—explained 25% of the variance in academic growth. Variation in growth was observed across racial/ethnic, gender, and family income groups. Students attending Catholic and private schools had the highest growth, whereas homeschooled students and students attending low-income public schools exhibited lower growth. Malleable factors associated with higher growth included earning higher grades in high school courses, taking elective high school courses in STEM areas, and taking advanced, accelerated, or honors courses. Students with Investigative and Conventional interests had higher growth.

8.3.3 Subgroup Differences in Growth

Academic Growth Patterns for English Language Learners and Students with Disabilities

Bassiri and Allen (2012) examined differences in growth for English learners (ELs) and students with disabilities (SWD). This study used longitudinal data on 103,725 students who took ACT Explore in 8th grade, ACT PLAN in 10th grade, and the ACT test in 11th or 12th grade. Across subject areas, the average ACT Explore, ACT PLAN , and ACT scores were highest for the reference group and lowest for the ELL and special education groups. Among SWD, the average scores across subject areas were lower for students with cognitive/learning disabilities than for students with physical disabilities. In some cases, the growth measures revealed a different pattern. For example, compared to the reference group, ELLs had consistently higher growth in mathematics and science between grade 8 and grade 10; SWD exhibited above-average growth in reading between grade 10 and grade 11/12, as did students with cognitive/learning disabilities in reading and science.

Most of the growth differences, while statistically significant, were small in magnitude. For example, students with a cognitive/learning disability increased scores 0.28 points less than the average in mathematics between grade 8 and grade 11/12, where the average gain for the reference group was 5.1 score points (16.5 on Explore to 21.6 on the ACT). The 0.28 growth difference was only about 5% (0.28 / 5.10) of the reference group's overall gain and so is small in magnitude. While many of the growth differences were not statistically significant or statistically significant but small in magnitude, some differences were more striking. Compared to their reference group peers, ELLs increased scores an average of 0.81 points less in English between grade 10 and 11/12 but grew nearly 0.75 points more in mathematics between grade 8 and 11/12; between grade 8 and 11/12, students in special education grew nearly a full point less in English, and students in the other disability group grew more than half a point less in mathematics.

Academic Growth Patterns of First-Generation College Students

Many college students are first-generation students, meaning neither parent attended college. First-generation college students tend to have lower college admission test scores and be less successful in completing their postsecondary programs than students whose parents went to college. Bassiri (2016a) investigated the extent to which gaps in their test scores might begin in middle school. This study used longitudinal data for approximately 282,000 students who took ACT Explore in 8th grade, ACT PLAN in 10th grade, and the ACT test in 11th or 12th grade. Four groups of students were identified according to their parents' highest grade level: no college experience (first-generation); some college experience, bachelor's degree, or graduate degree. The latter three groups were also referred to as non-first-generation. Following the same methodology as Bassiri and Allen (2012), residual gain scores were averaged for parents' educational level subgroups to form measures of aggregate growth for each subgroup and each grade-level period.

Across subject areas, the average ACT Explore, ACT PLAN, and ACT scores were positively associated with parents' education level. First-generation students and students whose parents had some college experience both had statistically significant ($p < .01$) and negative mean residual scores in all subject areas across all grade spans. Students whose parents had at least a bachelor's degree had statistically significant positive mean residual scores. Across all subject areas, it appeared that growth differences by parental education became more pronounced over time. That is, the mean residuals were larger in later grades, indicating that educational disparity by socioeconomic status increased over time. Future research should examine potential causes of the growth differences, such as low-income status, type of high school coursework, and high school grades.

8.4 Using Growth Models for Evaluation of Programs and School Effectiveness

8.4.1 Example of Program Evaluations

The federal government's Gaining Early Awareness and Readiness for Undergraduate Program (GEAR UP) was designed to increase the number of low-income students prepared to enter and succeed in postsecondary education. ACT, in partnership with the National Council for Community and Education Partnerships (NCCEP), conducted a research study to evaluate the effectiveness of GEAR UP programs with respect to students' academic readiness and college intent (ACT, 2007b).

In general, analyses suggested positive GEAR UP effects, though the effect sizes were generally small, and the statistically significant results were not consistent for the two cohorts studied. As stated in the report, the relatively small positive findings for the GEAR UP program may be underestimated due to limitations of the research design. For more details on the study's limitations and recommendations, see the full report (ACT, 2007b).

8.4.2 Measures of High School Effectiveness

In general, inferences about schools' effectiveness depend on the type of statistical model used to link student assessment results to schools.

Statistical Properties of Accountability Measures Based on ACT Assessments of College and Career Readiness

Allen, Bassiri, and Noble (2009) examined the statistical properties of different types of accountability models that use ACT test scores. The summary below focuses on accountability measures that attempt to measure the effects of high schools on ACT test scores. The results of the analyses indicated that different types of accountability measures can lead to different conclusions about a school's effectiveness. Because value-added models attempt to isolate the effects that schools have on student learning, they are less likely to be strongly related to school contextual factors. In most cases, estimated school effects do not differ significantly from the "average" school effect. This highlights the need for reporting the statistical uncertainty about estimates of schools' effects so that results can be properly interpreted.

Statistical Properties of School Value-Added Scores Based on Assessments of College and Career Readiness

Bassiri (2015) investigated methodological questions related to models for generating school effectiveness scores based on various ACT assessments across six different growth periods. The study found that value-added scores based on longer timeframes (e.g., grades 8–12) were more likely to distinguish school effects. The results underscored the association between prior academic achievement, particularly in the same subject but including off-subject scores, and future scores. Of school characteristics, prior mean academic achievement was positively related to the value-added measures, whereas school proportion of students eligible for free or reduced-price lunch and proportion of racial/ethnic minority students had negative relationships. Generally, compared to other school characteristics, class size and proportion of students tested had weaker associations with value-added measures. The importance of school characteristics varied by growth periods. When the ACT was the outcome variable, poverty level and class size tended to be more predictive of value-added scores. When ACT PLAN was the outcome variable, prior mean academic achievement tended to be more predictive. Value-added scores for low-poverty schools were higher than those obtained from high-poverty schools in all subject areas.

Relating Value-Added Measures of High School Effectiveness to Students' Enrollment and Success in College

Another study investigated the predictive strength of high school value-added measures on students' enrollment and success in college (Bassiri, 2016b). The study examined whether students from schools with higher value-added scores perform better in college. Measures of success in college included:

1. college enrollment in the fall after high school graduation,
2. grades in first-year college courses from four core content areas (English/language arts, mathematics, natural sciences, and social sciences), and
3. college retention from year one to year two.

The study found that value-added measures representing school effects on ACT scores had small but statistically significant relationships with college enrollment, college retention, and grades in first-year college courses in selected core content areas. The analyses controlled for student- and school-level characteristics that were also related to college success. The study revealed that the majority of the variance in college enrollment, retention, and in first-year course grades was due to students' characteristics; less of the variance was due to the characteristics of high schools or colleges. This was evidenced by their statistically significant but small intraclass correlation coefficient (ICC) estimates.

8.5 Retesting with the ACT

Many students take the ACT more than once. In 2015, 45% of ACT-tested high school students took multiple tests prior to graduating high school, up from 41% in 2009 (Harmston & Crouse, 2016). What are the typical score gains for students who retest with the ACT?

Lanier (1994) conducted an investigation of ACT Composite score gains and focused on how likely students were to obtain or exceed a specific ACT Composite score upon retesting, given their initial scores. In this investigation, the mean gain upon retesting was 0.8 scale score points. A follow-up study extended this research by describing typical ACT Composite score changes from first to second, second to third, and third to fourth testing, conditioned on first test score (Andrews & Ziomek, 1998). Approximately 95% of all students had a 70% to 80% chance of maintaining or increasing their score upon retesting. The percentage of examinees maintaining or increasing their score, as well as the average gain, decreased with each additional testing. The average ACT Composite score gain upon retesting was 0.75 points. Students with lower scores on previous tests had the greatest average gains, and those scoring near the maximum score of 36 had average score decreases.

Harmston and Crouse (2016) reexamined the trends associated with multiple testers, focusing on the number of times students took the ACT test and the time between tests. Score gains for multiple testers were highest for students who initially had low scores and for students who first tested in their sophomore year. Overall, ACT Composite score gains tended to be small for students who retested. Irrespective of these statistics, students should consider retesting if they believe their test scores do not accurately reflect their skills and knowledge. Test performance can be influenced by conditions prior to and during testing, including physical illness, temporary physical disabilities (e.g., broken arm), stress, or trauma.

Gains from the first to second ACT test have also been examined for over 772,000 students from the ACT-tested graduating class of 2013 who took the ACT two or more times (Camara & Allen, 2017). The results showed that 57% of students improved their ACT Composite score, 21% saw no change, and 22% saw a decrease in their ACT Composite score. For students with an initial ACT Composite score between 13 and 29, the typical gain in ACT Composite score from the first to second test was 1 point.

The studies described above examined ACT test-retest statistics descriptively. In a follow-up study of students from the 2018 high school graduating class who took the ACT two or more times, Harmston (2020) modeled students' chances of no ACT Composite score gain (including score drops) and gains of one, two, and three or more points on their second testing attempt as a function of student educational performance and behavioral attributes. The variables that were identified as having the strongest relationships with score gains included initial ACT Composite score; grade-level at time of first testing; time between two testing events; squared time between tests; interaction term between initial ACT Composite score and time between tests; HSGPA; indicator for whether planning to take physics in high school; indicator for whether planning to take calculus in high school; and indicator for whether planning to take one or more accelerated, honors, or advanced courses in high school. Results from this study were used to develop an ACT web application that enables users to calculate the likelihood of Composite score gains by student-specific criteria. For more details, see the full study.

Chapter 9
Other ACT Components

9.1 The ACT Interest Inventory

The primary purpose of the ACT® Interest Inventory is to stimulate and facilitate exploration of personally relevant educational and occupational (career) options. Given the important decisions and choices students must make as they navigate the transition from high school to college, exploration of self in relation to educational and occupational options is critical. Using their Interest Inventory results, students can explore programs of study and occupations in line with their interests.

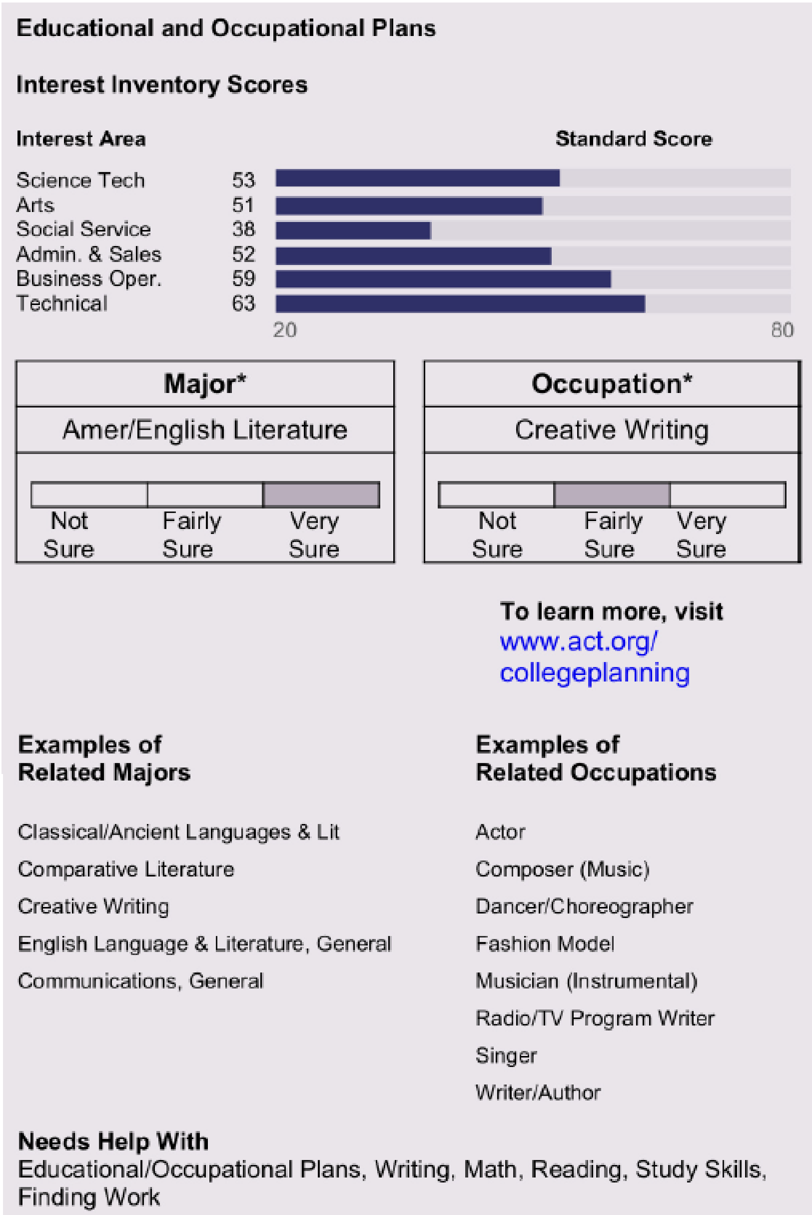
The ACT Interest Inventory consists of 72 items and provides scores on six scales that parallel Holland's (1997) six types of interests and occupations (see also Holland, Whitney, Cole, & Richards, 1969). Scale names (and parallel Holland types) are Science & Technology (Investigative), Arts (Artistic), Social Service (Social), Administration & Sales (Enterprising), Business Operations (Conventional), and Technical (Realistic). Each scale consists of common, everyday activities that are both familiar to students and relevant to work (e.g., study biology, help settle an argument between friends, sketch and draw pictures). The activities have been carefully chosen to assess basic work-relevant interests while minimizing the effects of sex-role connotations. Because males and females obtain similar distributions of scores, combined-sex norms are used to obtain sex-balanced scores. Readers seeking additional information about the ACT Interest Inventory are encouraged to consult *The ACT Interest Inventory Technical Manual* (ACT, 2023).

9.1.1 Reporting Procedures

High School Report

ACT Interest Inventory scores are reported as standard scores with a mean of 50 and a standard deviation of 10. The norms were based on a nationally representative sample of more than 250,000 12th-grade students from over 8,000 schools (for more information on the development of these norms, see ACT, 2023). These scores are made available to high school counselors, and they are combined with other information to provide education and career guidance (Figure 9.1). With knowledge of Holland's theory of career types (Holland, 1997), counselors may use ACT Interest Inventory scores to offer a clinical interpretation of the student's interests.

Figure 9.1. Example ACT Interest Inventory Results from the High School Report


Educational and Occupational Plans

Interest Inventory Scores

Interest Area	Standard Score
Science Tech	53
Arts	51
Social Service	38
Admin. & Sales	52
Business Oper.	59
Technical	63

20 80

Major*
Amer/English Literature
Not Sure · Fairly Sure · Very Sure

Occupation*
Creative Writing
Not Sure · Fairly Sure · Very Sure

To learn more, visit
www.act.org/
collegeplanning

Examples of Related Majors

Classical/Ancient Languages & Lit
Comparative Literature
Creative Writing
English Language & Literature, General
Communications, General

Examples of Related Occupations

Actor
Composer (Music)
Dancer/Choreographer
Fashion Model
Musician (Instrumental)
Radio/TV Program Writer
Singer
Writer/Author

Needs Help With
Educational/Occupational Plans, Writing, Math, Reading, Study Skills, Finding Work

Student Report

To facilitate educational and occupational exploration, results reported to students are expressed visually in work-world terms. Extensive research (much of it cited in Prediger, 1996) indicates that two orthogonal work-task dimensions (Data/Ideas and People/Things) underlie Holland's hexagonal model of interests and occupations (Holland, 1997; Holland, et al., 1969). Thus, a two-dimensional space can serve to display both a comprehensive set of occupations and the results of measured interests.

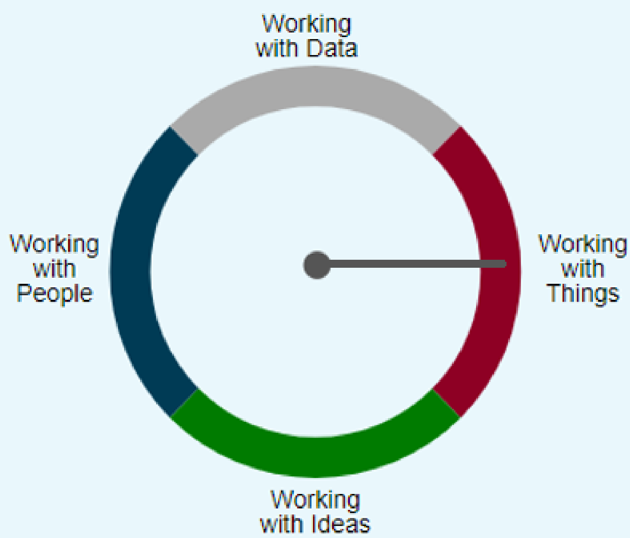
ACT Interest Inventory results are reported to students in two ways. First, it displays the results from the ACT Interest Inventory on the Career Connector. Second, it includes a short list of occupations that primarily involve the kinds of basic work tasks that the student prefers. The Career Connector is a two-dimensional figure with four compass points labeled Working with People, Data, Things, and Ideas (Figure 9.2; see ACT, 2023 for definitions). The Career Connector summarizes the pattern of scores on the six ACT Interest Inventory scales and visually displays it as one or two directions. For example, the point on a Career Connector may show that the student primarily enjoys activities involving ideas and people. The Career Connector is derived from the ACT Career Map, an empirically based system for summarizing basic similarities and differences between groups of occupations with respect to their relative involvement with people, data, things, and ideas. As described below, the Career Map serves as an interpretive bridge linking people to occupations by providing a visual display of personalized and actionable assessment results.

Figure 9.2. Example ACT Interest Inventory Results from the Student Report on MyACT

College and Career Planning

Interest-Work Preferences

Occupations differ widely in how much they involve working with four basic work tasks: Data, Ideas, People and Things. The results of your Interest Inventory point to the kinds of basic work tasks you prefer.



According to your results, you enjoy working with things. Here are a few examples of occupations involving this kind of work:

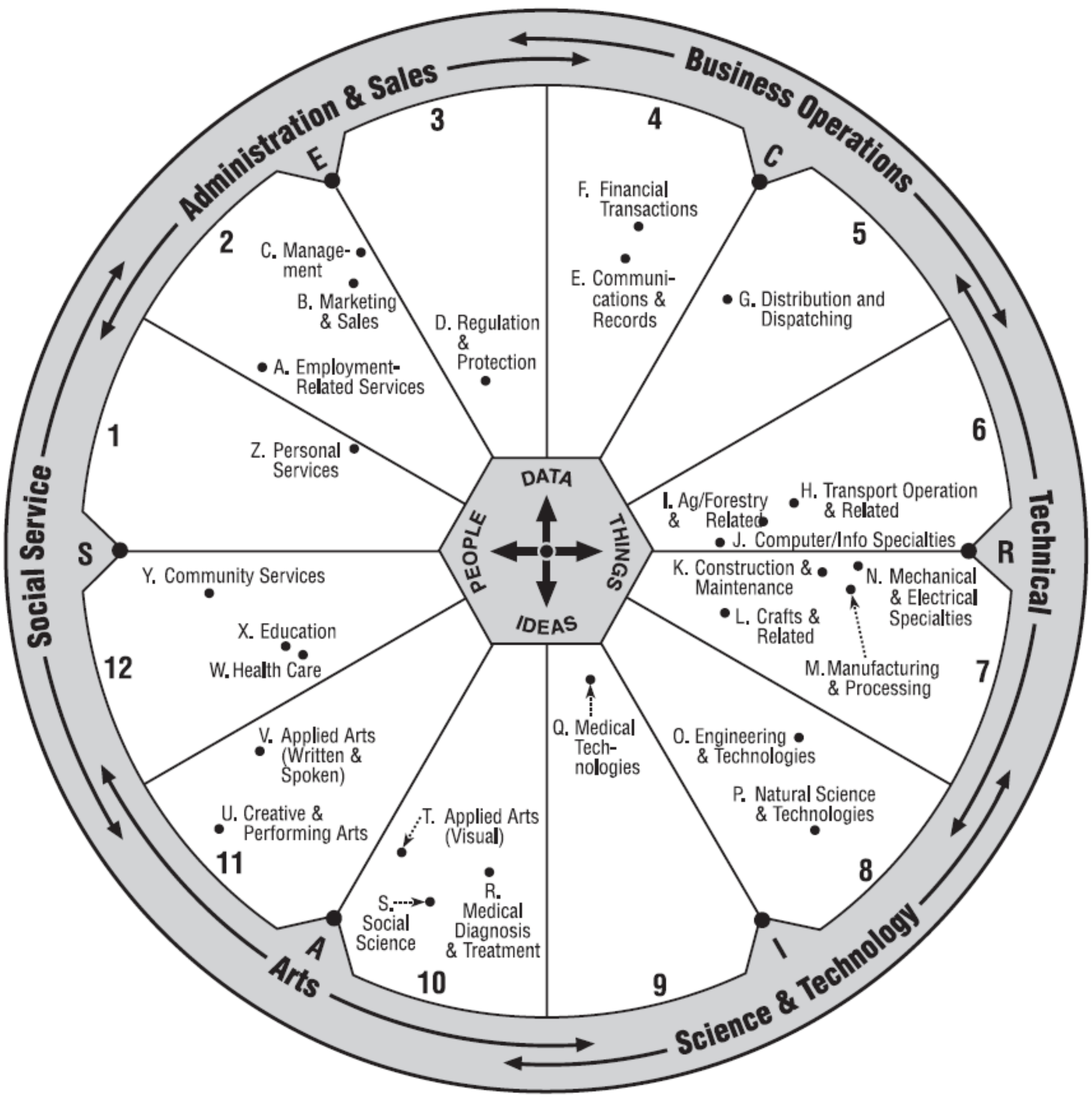
- Air Traffic Controller
- Broadcast Technician
- Computer Programmer
- Forester
- Machinist/Tool Programmer

For more details on how your interests align with your major choice and future career check out the plans tab.

Career Map

The ACT Career Map (Figure 9.3) provides a simple yet comprehensive overview of the world of work and provides a visual means for linking ACT Interest Inventory scores to career options. The 26 Career Areas (groups of related occupations) are located in 12 map “regions.” Career Areas are located on the Career Map according to the relative standing of their member occupations on the Data/Ideas and People/Things Work Task Dimensions. Career Area locations are based on extensive and diverse occupational data involving expert ratings, job analyses, and measured interests (ACT, 2023; Prediger & Swaney, 2004). The purpose of the work and work setting were also considered when the Career Areas were developed.

Figure 9.3. The ACT Career Map



Although care was taken to make each Career Area as homogeneous as possible, there is scatter across the occupations in each Career Area. The scatter could be reduced by the use of more Career Areas, but the Career Map was constructed for applied purposes and is not meant to provide a precise scientific statement. As can be seen in Figure 9.3, Career Area locations generally make good theoretical and common sense.

A student's pattern of ACT Interest Inventory scores is converted to map regions, and the Career Areas that align with the student's score pattern are reported, allowing for focused exploration of occupations that fit the student's interests. The method for converting scores to map regions is summarized in Appendix C of *The ACT Interest Inventory Technical Manual* (ACT, 2023).

9.1.2 Psychometrics

The ACT Interest Inventory Technical Manual (ACT, 2023), which presents a wide range of information about the inventory, includes the following topics:

- description of inventory items, scales, and interpretive aids
- development of items and norms
- reliability (internal consistency and test-retest stability)
- validity (convergent and discriminant evidence, item and scale structure, interest-environment fit, and success outcomes)

Internal consistency reliability coefficients for the six 12-item scales based on a grade 12 sample (N = 20,000) ranged from .84 to .91 (median = .87). Validity evidence is extensive, including discriminant validity evidence based on score profiles of 648 career groups (representing over 79,000 college major and occupation incumbents) and scale-structure evidence based on multiple samples (N = 60,000).

9.1.3 Interest-Major Fit

Interest-major fit is derived from two data elements collected during ACT test registration: the student's ACT Interest Inventory scores and the major the student plans to enter. Interest-major fit measures the strength of the relationship between the student's profile of ACT Interest Inventory scores and the profile of interests of students in the student's planned major. Interest profiles for each of the 294 majors on the ACT registration list are based on a large national sample of undergraduate students with a declared major and a GPA of at least 2.0. A student's major was determined in the third year for students in four-year colleges and in the second year for students in two-year colleges.

Interest-major fit scores range from 0 to 99. The higher the score, the better the interest-major fit. Using data from a large national sample, three levels of fit were established based on the empirical relationships between the interest-major fit scores and the proportion of students who persisted in their college major. In future score reports, level of interest-major fit will be displayed as shading of one of the three (Low, Medium, or High) sections of an Interest-Major Fit Bar.

Empirical evidence indicates that the fit between students' interests and their college majors is important to understanding and predicting student outcomes. Research involving the ACT

Interest Inventory suggests that if students' measured interests (i.e., patterns of interest scores) are similar to the interests of people in their chosen college majors, they will be more likely to persist in college (Tracey & Robbins, 2006; Allen & Robbins, 2008), remain in their majors (Allen & Robbins, 2008), and complete their college degree in a timely manner (Allen & Robbins, 2010). Even before students declare a major in college, fit between their interests and planned major is a good predictor of whether they will follow through on their college major plans (ACT, 2013c). The value of interest-major fit is not limited to the ACT Interest Inventory or to the outcomes listed above. A large-scale meta-analysis, which analyzed data over a 60-year period and included a range of outcome and interest measures (including the ACT Interest Inventory), found that interest-environment fit is related to persistence and performance in both academic and work settings (Nye, Su, Rounds, & Drasgow, 2012). Additional information on research involving the ACT Interest Inventory and interest-major fit is described in ACT (2023).

9.2 The MyACT Profile

Other student information is collected as part of registration for the ACT to broaden the information base for both students and colleges. The MyACT Profile information is collected and made available in a systematic form prior to college enrollment. The development of the MyACT Profile was influenced by the educational context in which it evolved. The information made available to students helps inform their educational choices, and the information made available to colleges provides a more comprehensive picture of the quality of education students received in high school.

The MyACT Profile contains several sections, each of which is discussed below. The items were developed by ACT staff with input from personnel representing a variety of postsecondary educational institutions. Items are revised from time to time to meet the changing needs of institutions for obtaining different types of data. The MyACT Profile questions and response options are located at https://www.act.org/content/act/en/myact-profile.html.

9.2.1 Demographic Information and Educational Opportunity Service

Students have the option to provide information on common demographics like age, race/ethnicity, gender, language, and parent/guardian education level in this section. ACT collects and reports this information in accordance with reporting guidelines issued by the U.S. Department of Education, which do not include collecting ethnic background information at a more detailed level. This information will be released to the colleges that receive student scores only if students request it. This section also provides students with the option to opt-in to the Educational Opportunity Service (EOS). This is a voluntary program for students to connect with colleges, scholarship agencies, and other educational opportunities seeking prospective students. Participating organizations have agreed to use student information only for the purpose of sharing information about their programs. ACT research shows that EOS participation benefits students by giving them greater access to information about colleges and programs of study, thereby expanding their possibilities for future success.

9.2.2 Your High School Resume

This section requests information about the student's high school performance (overall GPA, class rank), program of high school courses, the number of students in her or his graduating class, and the type of high school attended (public/private). If the student will graduate from a home school, this section asks how many years of high school homeschooling the student will have completed by graduation.

9.2.3 Your High School Courses & Grades

Most colleges, universities, and state agencies seek or require information from applicants on performance in a wide range of high school courses. To meet this need, ACT—in consultation with a representative group of personnel from postsecondary educational institutions—developed a list of 30 high school courses. In this section of the MyACT Profile, students are asked to report the grades they earned in these 30 courses, spanning six academic areas: English, mathematics, natural sciences, social studies, languages, and arts. ACT will calculate a grade point average (GPA) on an unweighted 4.0 scale based on the grade responses. High school GPAs based on self-reported grades are shown on the ACT College Report for English, mathematics, natural sciences, and social studies. For each subject area, students are asked the number of years of study expected by graduation. Students are also asked to indicate whether they have been enrolled in advanced placement, accelerated, or honors courses in any of five areas (English, mathematics, social studies, natural sciences, and foreign language). Because high school grades depend on both academic aptitude and personal characteristics such as persistence and study habits, these self-reports provide useful estimates of future academic achievement. Validity evidence for self-reported high school grades is discussed in Chapter 7.

9.2.4 College Plans

Questions about students' college plans are essential to the planning done by colleges. In this section, questions ask about students' plans for financing a college education. Students are asked to estimate their families' annual income and to indicate if they intend to apply for financial aid and/or to work while in college. Student responses to these questions can be useful to college financial aid officers.

Information about how students choose a college can be useful to personnel responsible for planning. This section also asks students questions about their enrollment plans (full-time or part-time), preferred type (public/private, two-year/four-year), size, and location of institution, along with preferred living accommodations. The student is asked to rank institutional preferences in order of importance.

With each new entering class, colleges must be prepared to provide individualized assistance to support the academic development of their students. Additional questions ask students to indicate their educational needs for improvement from a list of specific academic skills, including educational and occupational planning, writing, reading, study skills, and mathematics. By providing such information, students alert the college about their individual needs. This section also includes questions about student interest in college programs designed for enriched or accelerated academic work.

Entry into postsecondary education, as well as progress through such education, requires that students make important decisions and choices. Even tentative choices are important in that they provide a foundation for (and often place limits on) future finalized choices. The MyACT Profile questions provide opportunities for students to indicate information such as their intended college major, current occupational choice, and level of educational aspirations. This information is useful in evaluating the realism of student choices, as well as providing colleges with data that can be used for planning educational programs that meet the needs of all their students.

9.2.5 Extracurriculars

In this section, students are asked to indicate high school activities in which they have participated and activities in which they plan to participate during college. Students select from a list of activities such as athletics, drama, music, student government, student publications, and special-interest clubs. Information about students' current activities can be useful for connecting students to specific college opportunities. Information about the prospective plans of their incoming students is valuable for colleges seeking to develop appropriate extracurricular programs. From the student perspective, presenting their extracurricular activities and plans is another way of communicating their unique patterns of interests and skills.

References

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

ACT. (2005). *Courses count: Preparing students for postsecondary success*. Iowa City, IA: ACT.

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

ACT. (2007b). *Using Explore and Plan data to evaluate GEAR UP programs*. Iowa City, IA: ACT.

ACT. (2007c). *Case Study: Using EPAS to evaluate school-based intervention programs: GEAR UP*. Iowa City, IA: ACT.

ACT. (2009). *ACT National Curriculum Survey® 2009.* 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. (2013c). *College choice report part 3: Persistence and transfer*. Iowa City, IA: ACT.

ACT. (2015a). *ACT CAAP guide to successful general education outcomes assessment*. Iowa City, IA: ACT.

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

ACT. (2016a). *ACT Aspire Summative technical manual*. Iowa City, IA: ACT.

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

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

ACT. (2016d). *The ACT class profile national norms*. Iowa City, IA: ACT.

ACT. (2016e). *The ACT national high school profile report.* Iowa City, IA: ACT.

ACT. (2020a). *ACT Aspire technical manual.* Iowa City, IA: ACT.

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

ACT. (2020c). *PreACT 8/9 technical bulletin*. Iowa City, IA: ACT.

ACT. (2020d). *PreACT technical bulletin*. Iowa City, IA: ACT.

ACT. (2023). *The ACT Interest Inventory technical manual*. Iowa City, IA: ACT.

Allen, J. D. (2005). Grades as valid measures of academic achievement of classroom learning. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas*, *78*(5), 218–223.

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

Allen, J. (2015a). *Anticipated changes in ACT scores and participation rates with ACT statewide adoption.* Iowa City, IA: ACT.

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

Allen, J. (2015c). *Influence of achievement in core high school courses on ACT scores*. 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., Bassiri, D., & Noble, J. (2009). *Statistical properties of accountability measures based on ACT's educational planning and assessmen*t. 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.

Allen, J., Mattern, K., & Ndum, E. (2019). An empirically derived index of high school academic rigor. *Educational Measurement: Issues and Practices*, 38(1), 6–15.

Allen, J., & Radunzel, J. (2016). *Comparing high school grade point average (GPA) to first-year college GPA*. Iowa City, IA: ACT.

Allen, J., Radunzel, J., & Moore, J. L. (2017). *Evidence for standard setting: Probabilities of success in "Benchmark" college courses, by ACT test scores*. Iowa City, IA: ACT.

Allen, J., & Robbins, S. (2008). Prediction of college major persistence based on vocational interests, academic preparation, and first-year academic performance. *Research in Higher Education, 49*(1), 62–79.

Allen, J., & Robbins, S. (2010). Effects of interest-major congruence, motivation, and academic performance on timely degree attainment. *Journal of Counseling Psychology, 57*(1), 23–35.

Allen, J. & Sconing, J. (2005). *Using ACT Assessment scores to set benchmarks for college readiness*. 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.

Andrews, K. M., & Ziomek, R. L. (1998). *Score gains on retesting with the ACT assessment.* Iowa City, IA: ACT.

Bassiri, D., Allen, J. (2012). *Grade 8 to 12 academic growth patterns for English language learners and students with disabilities*. Iowa City, IA: ACT.

Bassiri, D. (2014). *The forgotten middle: Ensuring that all students are on target for college and career readiness before high school*. Iowa City, IA: ACT.

Bassiri, D. (2015). *Statistical properties of school value-added scores based on assessments of college readiness*. Iowa City, IA: ACT.

Bassiri, D. (2016a). *Academic growth patterns of first-generation college students in grades 8 to 12.* Iowa City, IA: ACT.

Bassiri, D. (2016b). *Are value-added measures of high school performance related to students' enrollment and college success?* Iowa City, IA: ACT.

Beatty, A. S., Walmsley, P. T., Sackett, P. R., Kuncel, N. R., & Koch, A. J. (2015). The reliability of college grades. *Educational Measurement: Issues and Practice*, *34*(4), 31–40.

Betebenner, D. W., Van Iwaarden, A., Domingue, B., and Shang, Y. (2014). SGP: An R package for the calculation and visualization of student growth percentiles & percentile growth trajectories. R package version 1.2-0.0. [Computer software] https://cran.r-project.org/web/packages/SGP/index.html

Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord and M. R. Novick (Eds.), *Statistical theories of mental test scores,* 397–479. Reading, MA: Addison-Wesley.

Brookhart, S. M. (1993). Teacher's grading practices: Meaning and values. *Journal of Educational Measurement, 30*(2), 123–142.

Brookhart, S. M. (2015). Graded achievement, tested achievement, and validity. *Educational Assessment, 20*(4), 268–296.

Camara, W., & Allen, J. (2017). *Does testing date impact student scores on the ACT?* Iowa City, IA: ACT.

Camara, W. J., Allen, J. M., & Moore, J. L. (2017). Empirically-based college and career readiness cut scores and performance standards. In K. L. McClarty, K. D. Mattern, & M. N. Gaertner (Eds.), *Preparing students for college and careers: Theory, measurement, and educational practice*. New York, NY: Routledge.

Camara, W. J., O'Connor, R., Mattern, K., & Hanson, M. A. (2015). *Beyond academics: A holistic framework for enhancing education and workplace success*. Iowa City, IA: ACT.

Casillas, A. (2013, October). *The other side of college readiness: Behaviors drive academic success*. Presentation at the Annual National Scholarships Providers Association Conference, Portland, OR.

Casillas, A., Way, J. D., & Burrus, J. (2015). Behavioral skills. In W. J. Camara, R. O'Connor, & K. Mattern (Eds.), *Beyond academics: A holistic framework for enhancing education and workplaces success* (pp. 25–38). Iowa City, IA: ACT.

CAST. (2011). *Universal design for learning guidelines version 2.0.* Wakefield, MA: West Virginia Department of Education. Retrieved from *https://udlguidelines.cast.org/binaries/content/assets/udlguidelines/udlg-v2-0/udlg_graphicorganizer_v2-0.pdf*.

Castellano, K. E., & Ho, A. D. (2013). *A practitioner's guide to growth models.* Washington, DC: CCSSO. Retrieved from *http://scholar.harvard.edu/files/andrewho/files/a_pracitioners_guide_to_growth_models.pdf*.

Castellano, K. E., & Ho, A. D. (2015). Practical differences among aggregate-level conditional status metrics: From median student growth percentiles to value-added models. *Journal of Educational and Behavioral Statistics*, *40*(1), 35–68.

Cizek, G. J., & Bunch, M. B. (2006). *Standard setting: A guide to establishing and evaluating performance standards on tests.* Thousand Oaks, CA: Sage Publications.

Clinedinst, M. (2015). *2014 state of college admission*. Alexandria, VA: National Association for College Admissions Counseling.

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.

Fedorchak, G. (2012). *Access by design—Implications for equity and excellence in education.* Draft paper prepared for the Smarter Balanced Assessment Consortium.

Fleiss, J. L., Levin, B., & Paik, M. C. (2003). *Statistical methods for rates and proportions* (3rd edition). Hoboken, NJ: John Wiley & Sons, Inc.

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.

Harmston, M. (2020). *Increasing your ACT Composite score: Should you test a second time?* Iowa City, IA: ACT.

Harmston, M., & Crouse, J. (2016). *Multiple testers: What do we know about them?* Iowa City, IA: ACT.

Harris, D. J., & Kolen, M. J. (1989). Examining group differences from a developmental perspective. In R. L. Brennan (Ed.), *Methodology used in scaling the ACT Assessment and PACT+* (pp. 75–86). Iowa City, IA: ACT.

Holland, J. L. (1997). *Making vocational choices: A theory of vocational personalities and work environments* (3rd ed.). Lutz, FL: PAR, Inc.

Holland, J., Whitney, D., Cole, N., & Richards, J. (1969). *An empirical occupational classification derived from a theory of personality and intended for practice and research*. Iowa City, IA: ACT.

Houston, M. (1993, April). *Accuracy of validity indices for course placement systems.* Paper presented at the annual meeting of the American Educational Research Association, Atlanta, GA.

Huh, N. R., & Huang, C. Y. (2016). *Examining the validity of ACT Composite scores and high school grade point average for predicting first-year college GPA of special-tested students*. Iowa City, IA: ACT.

Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings* (2nd ed.). Thousand Oaks, CA: Sage.

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.*

Koenker, R. (2005*). Quantile regression*. New York, NY: Cambridge University Press.

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.

Laing, J., & Farmer, M. (1984). *Use of the ACT assessment by examinees with disabilities*. Iowa City, IA: ACT.

Lanier, C. W. (1994). *ACT Composite scores of retested students*. Iowa City, IA: ACT.

Lawley, D. N. (1943). A note on Karl Pearson's selection formulae. *Proceedings of the Royal Society of Edinburgh, LX/I-Part, 1*(1), 28–30.

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.

Ling, J., & Radunzel, J. (2017). *Who is likely to graduate high school ready for college?* 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.

Lorah, J. & Ndum, E. (2013). *Trends in achievement gaps in first-year college courses for racial/ethnic, income, and gender subgroups: A 12-year study*. Iowa City, IA: ACT.

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.

Mattern, K., Burrus, J., Camara, W., O'Connor, R., Hansen, M. A., Gambrell, J., Casillas, A., & Bobek, B. (2014). *Broadening the definition of college and career readiness: A holistic approach*. Iowa City, IA: ACT.

Mattern, K. D., Patterson, B. F., Shaw, E. J., Kobrin, J. L., & Barbuti, S. M. (2008). *Differential validity and prediction of the SAT®*. (College Board Research Report No. 2008-4). New York, NY: College Board.

Mattern, K., Radunzel, J., Bertling, M., & Ho., A. D. (2018). How should colleges treat multiple admissions test scores? *Educational Measurement: Issues and Practice*, *37*(3), 11–23.

Mattern, K., Radunzel, J., & Westrick, P. (2015). *Development of STEM readiness benchmarks to assist educational and career decision making*. Iowa City, IA: ACT.

Mattern, K., Sanchez, E. I., & Ndum, E. (2017). Why do achievement measures underpredict female academic performance? *Educational Measurement: Issues and Practice. 36*(1), 47–57.

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.

Mislevy, R. J., Almond, R. G., & Lukas J. (2004). *A brief introduction to evidence-centered design.* (CRESST Technical Report 632). Los Angeles, CA: Center for the Study of Evaluation, CRESST, UCLA.

Mislevy, R. J., & Haertel, G. (2006). Implications of evidence-centered design for educational testing. *Educational Measurement: Issues and Practice, 25*(4), 6–20.

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

Moore, J., Huang, C., Huh, N., Li, T., & Camara, W. (2018). *Testing supports for English learners: A literature review and preliminary ACT research findings*. Iowa City, IA: ACT.

Moore, J. L., Li, T., & Lu, Y. (2020). *Reliability of English learners' test scores*. Iowa City, IA: ACT.

Moore, R., Sanchez, E., & San Pedro, S. (2019). *College entrance exams: How does test preparation affect retest scores?* Iowa City, IA: ACT.

McNeish, D. M., Radunzel, J., & Sanchez, E. I. (2015). *A multidimensional perspective of college readiness: Relating student and school characteristics to performance on the ACT®*. Iowa City, IA: ACT.

National Center for Education Statistics. (2021). *Digest of education statistics 2019* (NCES 2021-009). Washington, DC: U.S. Department of Education. Retrieved from *https://nces.ed.gov/pubs2021/2021009.pdf.*

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.

Ndum, E. A., Allen, J., Way, J. D., & Casillas, A. (2015). *Effects of psychosocial factors on gender differences in first-year college English and mathematics grades*. Iowa City, IA: ACT.

Ndum, E., & Mattern, K. (2016). *Gender gaps in STEM preparation and interests over time*. Iowa City, IA: ACT.

Ndum, E., Radunzel, J., & Westrick, P. (2016). *ACT performance for students with disabilities who tested with accommodations in 2014.* Iowa City, IA: ACT.

Noble, J., Crouse, J., Sawyer, R., & Gillespie, M. (1992, April). *Ethnic/gender bias and the differential performance hypothesis: Implications for performance on the ACT assessment.* Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Noble, J., Crouse, J., & Schulz, M. (1995, April). *Differential prediction/impact in course placement for ethnic and gender groups.* Paper presented at the annual meeting of the National Council on Measurement in Education, San Francisco, CA.

Noble, J., Crouse, J., & Schultz, M. (1996). *Differential prediction/impact in course placement for ethnic and gender groups*. Iowa City, IA: ACT.

Noble, J., Davenport, M., & Sawyer, R. (2001, April). *Relationships between noncognitive characteristics, high school course work and grades, and performance on a college admissions test*. A paper presented at the annual meeting of the American Education Research Association, Seattle, WA.

Noble, J., Davenport, M., Schiel, J., & Pommerich, M. (1999a). *Relationships between the noncognitive characteristics, high school course work and grades, and test scores for ACT-tested students*. Iowa City, IA: ACT.

Noble, J., Davenport, M., Schiel, J., & Pommerich, M. (1999b). *High school academic and noncognitive variables related to the ACT scores of racial/ethnic and gender groups*. Iowa City, IA: ACT.

Noble, J., & McNabb, T. (1989). *Differential course work in high school: Implications for performance on the ACT Assessment*. Iowa City, IA: ACT.

Noble, J., Roberts, B., & Sawyer, R. L, (2006*). Student achievement, behavior, perceptions and other factors affecting ACT scores*. Iowa City, IA: ACT.

Noble, J., & Sawyer, R. (2013). *A study of the effectiveness of developmental courses for improving success in college*. Iowa City, IA: ACT.

Noble, J., & Sawyer, R. (2002). *Predicting different levels of academic success in college using high school GPA and ACT Composite score*. Iowa City, IA: ACT.

Noble, J. P., Schiel, J. L., & Sawyer, R. L. (2004). Assessment and college course placement: Matching students with appropriate instruction. In J. E. Wall & G. R. Walz (Eds.), *Measuring up: Assessment issues for teachers, counselors, and administrators* (pp. 297–311). Greensboro, NC: ERIC Counseling & Student Services Clearinghouse and the National Board of Certified Counselors.

Noble, J., & Schnelker, D. L. (2007). *Using hierarchical modeling to examine course work and ACT score relationships across high schools*. Iowa City, IA: ACT.

Nye, C., Su, R., Rounds, J., & Drasgow, F. (2012). Vocational interests and performance: A quantitative summary of over 60 years of research. *Perspectives on Psychological Science, 7*(4), 384–403.

Patterson, B. F., & Mattern, K. D. (2012). *Validity of the SAT for predicting first-year grades: 2009 SAT validity sample*. New York, NY: College Board.

Payne, J. S., & Allen, J. (2019). *An early look at ACT Academy usage and effectiveness*. Iowa City, IA: ACT.

Pilcher, J. K. (1994). The value-driven meaning of grades. *Educational Assessment, 2*(1), 69–88.

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

Poropat, A. E. (2009). A meta-analysis of the five-factor model of personality and academic performance. *Psychological Bulletin*, *135*(2), 322–338.

Prediger, D. (1996). Alternative dimensions for the Tracey-Rounds interest sphere. *Journal of Vocational Behavior, 48*(1), 59–67.

Prediger, D., & Swaney, K. (2004). Work task dimensions underlying the world of work: Research results for diverse occupational databases. *Journal of Career Assessment, 12*(4), 440–459.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*. Thousand Oaks, CA: Sage Publications.

Radunzel, J. (2012, October). *Where are 2003 high school graduates seven years later?* Paper presented at the Eighth Annual National Symposium on Student Retention, New Orleans, LA. Retrieved from *http://www.act.org/content/dam/act/unsecured/documents/WhereDidHighSchoolGraduatesGo.pdf*.

Radunzel, J. (2019). *Relating ACT writing scores to first-year English composition grades*. Iowa City, IA: ACT.

Radunzel, J., & Allen, J. (2020). *Predicting success on Advanced Placement exams using ACT Aspire, PreACT, and ACT test scores.* 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., Mattern, K., & Westrick, P. (2016). *The role of academic preparation and interest on STEM success*. Iowa City, IA: ACT.

Radunzel, J., & Noble, J. (2012). *Predicting long-term college success through degree completion using ACT Composite score, ACT Benchmarks, and high school grade point average*. Iowa City, IA: ACT.

Radunzel, J., & Noble, J. (2013). *Differential effects on student demographic groups of using ACT College Readiness Assessment Composite score, ACT benchmarks, and high school grade point average for predicting long-term college success through degree completion*. Iowa City, IA: ACT.

Radunzel, J., Noble, J., & Wheeler, S. (2014). *Dual-credit/dual-enrollment coursework and long-term college success in Texas.* 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.

Sanchez, E. I. (2013). *Differential effects of using ACT college readiness assessment scores and high school GPA to predict first-year college GPA among racial/ethnic, gender, and income groups*. Iowa City, IA: ACT.

Sanchez, E. (2018). *ACT Online Prep helps students on the ACT*. Iowa City, IA: ACT.

Sanchez, E. (2019). *Can using ACT Online Prep improve score gains?* Iowa City, IA: ACT.

Sanchez, E. (2020). *ACT Online Prep: An effective way to raise ACT scores*. Iowa City, IA: ACT.

Sanchez, E. I., & Buddin, R. (2016). *How accurate are self-reported high school courses, course grades, and grade point average?* Iowa City, IA: ACT.

Sawyer, R. L. (1985). *Using demographic information in predicting college freshman grades*. 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.

Sawyer, R. L. (2008). *Benefits of additional high school course work and improved course performance in preparing students for college.* Iowa City, IA: ACT.

Sawyer, R. (2010). *Usefulness of high school average and ACT scores in making college admission decisions*. Iowa City, IA: ACT.

Sawyer, R. (2013a). Beyond correlations: Usefulness of high school GPA and test scores in making college admissions decisions. *Applied Measurement in Education*, *26*(2), 89–112.

Sawyer, R. (2013b). *Interpreting changes over time in high school average, ACT College Readiness Assessment Composite scores, and ACT College Readiness Benchmark attainment rates*. Iowa City, IA: ACT.

Sawyer, R., Laing, J., & Houston, M. (1988). *Accuracy of self-reported high school courses and grades of college-bound students* (ACT Research Report No. 88-1). Iowa City, IA: American College Testing Program.

Schiel, J. (1998). *Academic benefits in high school of an intensive summer program for academically talented seventh graders* (ACT Research Report No. 98-4). Iowa City, IA: ACT.

Schiel, J. L. (2020). *Test-related behaviors and performance on the ACT.* Iowa City, IA: ACT.

Schiel, J., & Harmston, M. (2000). *Validating two-stage course placement systems when data are truncated*. Iowa City, IA: ACT.

Schiel, J., & Noble, J. (1992). *The effects of data truncation on estimated validity indices for course placement.* Iowa City, IA: ACT.

Schiel, J., Pommerich, M., & Noble, J. P. (1996). *Factors associated with longitudinal educational achievement, as measured by PLAN and ACT Assessment scores.* Iowa City, IA: ACT.

Schiel, J. L., & Valiga, M. J. (2014a). *How ACT scores change with test preparation.* Iowa City, IA: ACT.

Schiel, J. L., & Valiga, M. J. (2014b). *How the amount of time spent on test preparation relates to ACT scores.* Iowa City, IA: ACT.

Schiel, J. L., & Valiga, M. J. (2014c*). How particular outcomes of test preparation relate to ACT scores.* Iowa City, IA: ACT.

Scholes, R. J., & Lain, M. M. (1997, March). *The effects of test preparation activities on ACT Assessment scores.* A paper presented at the annual meeting of the American Educational Research Association, Chicago, Illinois.

Scholes, R. J., & McCoy, T. R. (1998, April). *The effects of type, length, and content of test preparation activities on ACT Assessment scores.* A paper presented at the annual meeting of the American Educational Research Association, San Diego, California.

Sinharay, S., Haberman, S. J., & Lee, Y.-H. (2011). When does scale anchoring work? A Case Study. *Journal of Educational Measurement*, *48*(1), 61–80.

Snyder, T. D., & Dillow, S. A. (2013). *Digest of education statistics 2012* (NCES 2014-015). National Center for Education Statistics, Institution of Education Sciences, US Department of Education, Washington, DC.

Sparks, D., & Malkus, N. (2013). *First-year undergraduate remedial coursetaking: 1999-2000, 2003-04, 2007-08* (Statistics in Brief, NCES 2013-013). Washington, DC*:* National Center for Education Statistics.

Spray, J. A. (1989). *Performance of three conditional DIF statistics in detecting differential item functioning on simulated tests.* 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.*

Stiggins, R. J., Frisbie, D. A., & Griswold, P. A. (1989). Inside high school grading practices: Building a research agenda. *Educational Measurement: Issues and Practice, 8*(2), 5–14.

Stocking, M. L., & Lord, F. M. (1983). Developing a common metric in item response theory. *Applied Psychological Measurement*, *7*(2), 201–210.

Tracey, T., & Robbins, S. (2006). The interest-major congruence and college success relation: A longitudinal study. *Journal of Vocational Behavior*, *69*(1), 64–89.

United States Department of Education. (1994). *What do student grades mean? Differences across schools* (Office of Research Report 94-3401). Washington, DC: Office of Educational Research and Improvement.

Valiga, M. (1986). *Accuracy of self-reported high school course and grade information*. Iowa City, IA: ACT.

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: *https://www.maine.gov/doe/sites/maine.gov.doe/files/inline-files/dok.pdf.*

Westrick, P. (2016). *The joint use of ACT® scores and high school grade point average for predicting success at community colleges*. Iowa City, IA: ACT.

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

Westrick, P., & Allen, J. (2014). *Validity evidence for ACT Compass placement tests*. Iowa City, IA: ACT.

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.

Westrick, P., Schmidt, F., Le, H., Robbins, S., & Radunzel, J. (2021). The road to retention passes through first year academic performance: A meta-analytic path analysis of academic performance and persistence. *Educational Assessment*, *26*(1), 35–51.

Whitney, D. R. (1989). Educational admissions and placement. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 515–525). New York, NY: American Council on Education, and Macmillan.

Willingham, W. W., Pollack, J. M., & Lewis, C. (2002). Grades and test scores: Accounting for observed differences. *Journal of Educational Measurement*, *39*(1), 1–37.

Woodruff, D. J., & Ziomek, R. L. (2004a). *High school grade inflation from 1991 to 2003*. Iowa City, IA: ACT.

Woodruff, D. J., & Ziomek, R. L. (2004b). *Differential grading standards among high schools*. Iowa City, IA: ACT.

Zhang, Q., & Sanchez, E. I. (2013). *High school grade inflation from 2004 to 2011*. Iowa City, IA: ACT.

Zimowski, M., Muraki, E., Mislevy, R. J., & Bock, R. D. (2003). BILOG-MG (Version 3.0) [Computer software]. Chicago, IL: Scientific Software International.

Ziomek, R. L., & Andrews, K. M. (1996). *Predicting the college grade point averages of special-tested students from their ACT assessment scores and high school grades*. Iowa City, IA: ACT.

Ziomek, R. L., & Svec, J. C. (1995). *High school grades and achievement: Evidence of grade inflation*. Iowa City, IA: ACT.