ACT® Technical Manual

The ACT®
Technical Manual

Fall 2019 Version 3

Table of Contents

List of Tables

List of Figures

Preface

The ACT® Technical Manual contains detailed technical information about the ACT test. The principal purpose of the manual is to document technical characteristics of the ACT test in light of its intended purpose. *The ACT Technical Manual* documents the collection of validity evidence that supports appropriate interpretations of test scores and describes various content and psychometric aspects of the ACT. Multiple test design and development processes are articulated documenting how ACT attends to building 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 ongoing and continuous improvement and research intended to assure that the program remains psychometrically sound.

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.

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 programs (all leading up to the ACT®) have provided students, educators, and policy makers 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 academic skills they will need to perform college-level work. The required academic skills can be assessed most directly by reproducing, as faithfully as possible, the complexity of college-level work. Therefore, the 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 have a direct relationship to the students' educational progress in curriculum-related areas and possess a meaning that is readily grasped by students, parents, and educators.

The ACT constructs in the subject tests are supported by multiple sources of validity evidence. ACT has multiple years of longitudinal statistical evidence backing the strong relationship between student performance on the subject tests and student performance in entry level courses in those subjects. More recent methodologies such as cognitive labs have served to further confirm this evidence.

The advantage of tests of educational achievement over other types of tests for use in the transition from high school to college and the workplace becomes evident when their use is considered in the context of the educational system. Because tests of educational achievement measure many of the same skills that are taught in high school, the best preparation for tests of educational achievement is high school coursework. Long-term learning in school, rather than short-term cramming and coaching, becomes the obvious best form of test preparation. Thus, tests of educational achievement tend to serve as motivators by sending students a clear message that high test scores are not simply a matter of innate ability, but reflect a level of achievement that has been earned as a result of hard work.

Because the ACT stresses such general concerns as the complexity of college-level work and the integration of knowledge from a variety of sources, students may be influenced to acquire skills necessary to handle these concerns. In this way, the ACT may serve to aid high schools in developing the critical thinking skills in their students important for success in college and later life.

The tests of the ACT therefore are designed not only to accurately reflect educational goals that are widely accepted and judged by educators to be important, but also to give educational considerations, rather than statistical and empirical techniques, paramount importance.

1.2 Overview and Purpose of the ACT

The ACT emphasizes meaningful outcomes, placing concepts of readiness and competency at the forefront of 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, reading, mathematics, and science—along with an optional writing test. As part of registration for the test, ACT also collects information about students' high school courses and grades, educational and career aspirations, extracurricular activities, and special educational needs.

ACT data are used for many purposes. Students use their results to plan for further education and explore careers based on their own skills, interests, and aspirations. When they know what colleges expect, in terms they can understand, students can take ownership and control of their information, and they can use it to help make a smooth transition to postsecondary education or training. High schools use ACT data in academic advising and counseling, evaluation studies, accreditation documentation, and public relations. Colleges use ACT results for admissions and course placement. States use the ACT as part of their statewide assessment and accountability. Many of the agencies that provide scholarships, loans, and other types of financial assistance to students tie such assistance to students' academic qualifications, as measured by ACT scores. Many state and national agencies also use ACT data to identify talented students and award scholarships.

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)—statements that describe students' performance in terms of the knowledge and skills they have acquired. 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 and Career Readiness Benchmarks are the minimum score associated with a high likelihood of post-secondary success in each content area. Together, the ACT College and Career Readiness Benchmarks and the CCRS provide students specific insight to succeed in college and career. Chapter 8 gives details about the College and Career Readiness Standards and the College and Career Readiness Benchmarks.

Table 1.1 summarizes the assessment components.

Table 1.1. Components of the ACT

Component	
Career and Education Planning	Interest Inventory Course Taking and Grades Student Profile
Objective Assessments	English Mathematics Reading Science Writing (optional)
Instructional Support	College and Career Readiness Standards
Test Preparation (free)	ACT® Academy™ ACT Question of the Day Preparing for the ACT Online Familiarity Assessment Alternate Assessment Formats
Evaluation	Summary Reports

ACT Score Reports

The ACT student score reports present data visually that is engaging and provides additional content, including reporting category scores. Reporting categories are reported for each subject test (English, mathematics, reading, and science). The reporting categories are based on the ACT College and Career Readiness Standards.

Each reporting category is based on a subset of items in the subject test. For each reporting category, the score report shows the following:

- Total points possible
- Total points achieved
- Percent correct
- The ACT Readiness Range: this allows students to compare their performance in each reporting category to the performance of students who have met the ACT College and Career Readiness Benchmark in that subject.

Additionally, the ACT provides a Work Readiness indicator that informs students how they are progressing toward earning the ACT® WorkKeys® National Career Readiness Certificate® (NCRC®). This additional feature assists in determining whether students are on the pathway to being college and career ready.

The report is accompanied by a booklet, *Using Your ACT Results*, which provides interpretive information about the test results, describes ACT services and policies, and tells examinees how to contact ACT for further information.

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

In creating the ACT, a theory of action (TOA) was employed that integrates content validity (academic research, curriculum information, and academic standards) with predictive validity (empirical data). The TOA begins by answering fundamental questions about the purpose, users, uses, benefits, claims, interpretations, and outcomes.

Intended Purpose. The primary purpose of the ACT is to measure students' level of college and career readiness in key core academic content areas. The test is a comprehensive system of data collection, processing, and reporting designed to help high school students develop postsecondary educational plans and to help postsecondary educational institutions meet the needs of their students.

The ACT provides overall score and subject test scores for each of the four subject tests and the optional writing test. The test also provides a measure of a students' STEM skills (by combining mathematics and science scores), an Understanding of Complex Texts (UCT) indicator, and a combined ELA score (by combining English, reading, and writing scores for students who take the writing test). The test also provides information at a more detailed level (i.e., reporting categories). With the exception of the writing and UCT scores, each score is reported on a scale that ranges from 1 to 36. The writing score is reported on a 2–12 scale. The UCT score is reported using three levels: Below Proficient, Proficient, and Above Proficient.

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 interpretations are used for many intended purposes. Students use their results to plan for further education and explore careers based on their own 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 for admission and course placement decisions. States use the ACT as part of their statewide assessments to measure students' educational achievement and to monitor educational improvement and achievement gaps over time. Many private, state, and national agencies that provide scholarships, loans, and other types of financial assistance to students tie such assistance to students' academic qualifications, as measured by ACT test scores.

Intended Benefits. The intended benefits of using the ACT test include:

- allowing students to demonstrate their knowledge and skills gained throughout educational course work in core content areas of English, mathematics, reading, and science;
- providing students with a profile of their relative strengths and weaknesses in the subject areas assessed by the test informing what an examinee knows 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;

- assisting students to better prepare for college and careers;
- providing indicators as to whether a student is likely ready for college-level course work or a work training program (based on the College and Career Readiness Benchmarks and the Progress Toward the ACT National Career Readiness certificate); and
- providing colleges and talent identification and scholarship agencies with information about students' level of achievement in the subject areas assessed by the test.

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

- The ACT measures the academic knowledge and skills that are acquired in high school and are important for being ready for college-level course work 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 and course placement 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 the school administers the ACT to all of their students.
- The ACT includes the ACT Interest Inventory, which is based on documented evidence of validity for career planning purposes, 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 the ACT test scores, provide a more holistic picture of the student's educational development and career-relevant motivations.

Intended Outcomes. The intended outcomes from using the results of the ACT in conjunction with other academic and non-academic measures include, but are not limited to helping:

- students, parents, and educators to identify academic knowledge and skills where students might benefit from additional instruction and supports while still in high school to better prepare for college and career and avoid needing to take remedial or developmental courses in their first year of college;
- students to expand their educational and occupational exploration and opportunities beyond options initially considered based on students' academic strengths and weaknesses and interests measured from the ACT Interest Inventory (ACT, 2009) or through ACT's Educational Opportunity Service (Moore & Cruce, 2017);
- schools and districts to raise college awareness and exposure among all students when state or district testing of the ACT is used;
- schools and districts to evaluate student growth and identify gaps in educational achievement in order to better inform school programs that are effective in preparing all students for college and career readiness;
- postsecondary institutions to select students for admissions who are likely to enroll at their institution and once enrolled, likely to succeed in their college courses and persist and complete a college degree at their institution;

- postsecondary institutions to place students in first-year college courses in which they are most likely to be successful; and
- postsecondary institutions to identify students early on who are most likely to struggle academically, be at risk of dropping out of college, and 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

Artifacts of the ACT test emerge from an evidence-based research and data collection process to ensure that items and test forms are eliciting the intended evidence to support the claims made by the ACT. For example, content and item specifications and test blueprints influence the technical quality and output of test items and forms. These artifacts are 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®
- Survey of standards frameworks—including, but not limited to the ACT College and Career Readiness Standards (CCRS), the Next Generation Science Standards (NGSS), and other college and career readiness standards

ACT's National Curriculum Survey provides empirical validation evidence related to the content of the tests. The most recent survey was released in 2016 and included responses from thousands of educators from K–12 to college instructors in ELA, mathematics, science, and reading.

The ACT National Curriculum Survey includes workforce supervisors and employees to provide evidence relating to the skills and knowledge essential for career readiness. Results are reviewed by SMEs and used to identify the most critical skills and knowledge required for college and career readiness.

The validation argument is further supported with criterion-related longitudinal evidence tracking students who complete the ACT to subsequent performance in colleges (two-year and four-year, by major) and career training programs.

While hundreds of discrete skills and knowledge items can be identified by SMEs as relevant to high school curriculum, not every skill and knowledge is essential for postsecondary success, nor will every skill differentiate students. Some skills, which may be essential for success may be attained by more than 95 percent of students continuing on with postsecondary education and including items that measure these skills on a test only increases test length without any contribution to measuring college success and prediction outcomes.

Similarly, our 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 test development prioritizes the skills and knowledge most essential to success based on empirical data and that generalize across different institutions, academic programs, and majors.

SME judgment is an important component of the development process, but unlike many other programs, it is influenced by empirical data and research bases from actual test takers who then proceed to postsecondary education.

The first step in building this architecture is to synthesize research on high-value skill targets—the skill targets that can be shown to offer the most useful evidence for college and career readiness. This evidence is achieved by organizing units of knowledge and skills into levels.

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

Test blueprints specify constraints that serve to control various factors, including, but not limited to, content coverage, item difficulty, cognitive complexity, reading load, and item latency. Test forms are then administered and student performance data are collected.

The following diagram helps to illustrate how a validity argument is composed of multiple sources of research, empirical data, and other forms of evidence. Content validity is shown coming from the research base. Predictive validity information flows in primarily from the ACT and, to a lesser extent, the ACT WorkKeys® assessments. Both channels of information feed into the knowledge and skills needed to perform well on the ACT, thus supporting an iterative model of refinement that serves the common goal of informing whether a student is "on track" for college and career readiness.

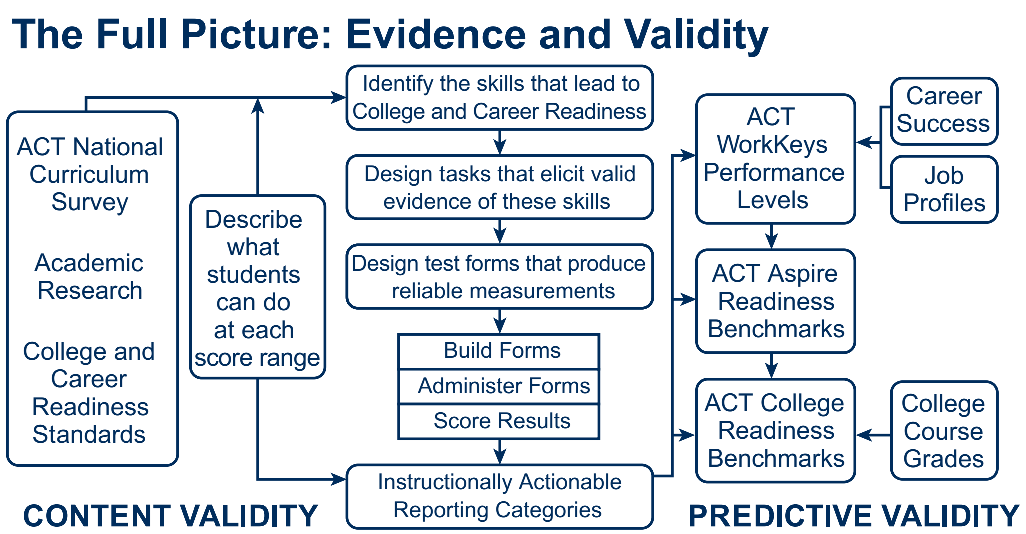


Figure 1.1 The full picture: evidence and validity

1.5 Code of Fair Testing Practices in Education and Code of Professional Responsibilities in Educational Measurement

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, 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.

ACT's endorsement of the *Code* represents a commitment to vigorously safeguard the rights of individuals participating in its testing programs. ACT employs an ongoing review process whereby each of its testing programs is routinely reviewed to ensure that it upholds the standards set forth in the *Code* for appropriate test development practice and test use.

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.

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 sixteen and twenty.

Self-reported data describing the ACT examinee population for the 2018 high school graduating class are presented in Table 1.2. These data are based on the 1,914,817 students who graduated in the spring of 2018 and who took the ACT either during their sophomore, junior, or senior year in high school. For students who took the test two or more times, the most current test score is used.

Historically, ACT has advised students to take the ACT after they have completed a substantial portion of the coursework covered by its tests. Given the curriculum of most secondary schools and the course of study followed by the majority of the students, this point is usually reached by spring of the junior year. However, this varies from student to student and with the four academic areas measured by the ACT.

Table 1.2 Demographic Characteristics of the 2018 ACT-Tested High School Graduating Class

Demographic	%[a]
Gender	
Female	52
Male	47
No response	2
Grade Level When Tested	
Senior	46
Junior	53
Other	1
No response	<1
Racial-Ethnic Background	
African American/Black	13
White	56
American Indian/Alaska Native	1
Hispanic/Latino	15
Asian	4
Native Hawaiian/Other Pacific Islander	<1
Two or more races	4
Prefer no response/blank	6

[a]Due to rounding, some columns may not add to exactly 100%.

1.7 Test Preparation

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

- **ACT Academy.** Provides teachers with powerful tools to support every unique student in the classroom. Teachers can directly assign additional practice—with a single click—to all of their students, a group of students, or multiple classes based on student performance and learning gaps. In addition, ACT Academy provides teachers with powerful analytics about individuals and classrooms to drive the teaching and learning cycle and improve outcomes for each student.
- **ACT Question of the Day.** We post a daily test question 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 prior week.
- **Preparing for the ACT or Preparación Para el Examen de ACT.** Includes a full-length practice test, test-taking strategies, and what to expect on test day. This publication is available in English and Spanish and is available as a free download by teachers, students, parents, and others: www.act.org/content/dam/act/unsecured/documents/Preparing-for-the-ACT.pdf, www.act.org/content/dam/act/unsecured/documents/Preparing-for-the-ACT-Spanish.pdf.
- **Online Familiarity Assessment.** A full-length practice test available in our simulated online testing experience. Students may also access both timed and untimed practice tests for each ACT subject. Students may sign into each of the subject 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 scripts are available at no cost to the school and contain a full-length practice test.

References

ACT. (2009). *The ACT interest inventory technical manual.* Iowa City, IA: ACT. Retrieved from https://www.act.org/content/dam/act/unsecured/documents/ACTInterestInventoryTechnicalManual.pdf

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

Joint Committee on Testing Practices. (2004). *Code of Fair Testing Practices in Education.* Washington, DC: Author.

Moore, J. L., & Cruce, T. (2017). *Does opting into a search service provide benefits to students?* (ACT Working Paper 2017-3). Iowa City, IA: ACT.

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.

Chapter 2

The ACT® Test Development

2.1 Overview

This chapter describes ACT's test development process—including item and form development procedures, and the ACT writing test prompt construction process. A brief overview of the National Curriculum Survey, content and bias review process and statistical criteria for selecting operational items and form assembly is also included. Lastly, information includes a high-level description of the ACT scoring procedures, including descriptions of additional scores and indicators.

2.2 Description of the ACT Tests

The ACT contains four tests—English, mathematics, reading, and science—and an optional writing test. These tests measure the most important content, skills, and concepts taught in high school and are 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 feedback directly from current high school and postsecondary teachers (via the ACT® National Curriculum Survey®, and panels) as well as student data from the ACT and from actual postsecondary performance in courses. These empirical data are used to continually verify the knowledge and skills required for postsecondary and career success and being measured by the ACT.

2.3 The ACT National Curriculum Survey

The ACT National Curriculum Survey is a one-of-a-kind nationwide survey of educational practices and college and career readiness expectations conducted by ACT every few years. ACT surveys thousands of K–12 teachers and college instructors in English/writing, mathematics, reading, and science—and, for the first time in 2016, 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 for college and career readiness.

Also for the first time in 2016, questions were included about which skills from ACT‘s Holistic Framework™—a holistic, research-based framework that integrates the knowledge and skills that empower people to achieve education and career success—are most integral to college and career success. These included behavioral skills, education and career navigation skills, and dimensions such as core academic skills and cross-cutting capabilities.

ACT uses the results of the ACT National Curriculum Survey to guide the development of ACT assessment solutions, including the ACT test, ACT Aspire®, and ACT WorkKeys®. ACT conducts the survey to ensure that its assessments are measuring the current 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 because ACT data can help education and workforce stakeholders make more informed decisions about the skills needed to be successful in postsecondary education and the workplace.

An Integrated Framework for Education and Career Success

The ACT National Curriculum Survey is an essential tool in ACT's commitment to ensuring not only that the assessments are valid and relevant on a continuing basis, but also that they provide information enabling students and workers to be fully ready to embark successfully on rewarding college and career journeys.

The Purpose of the ACT National Curriculum Survey

The ACT National Curriculum Survey is a crucial step in the process used to build and regularly update a valid suite of ACT assessments that is empirically aligned to college readiness standards. The survey directly informs the test blueprint for the assessments. 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 actual course performance, ACT answers a second critical question: Does the test accurately and reliably predict performance? Constant monitoring allows ACT to ensure that the answer to both questions is "yes."

ACT begins development of its assessments by using the dual validity loop and the ACT National Curriculum Survey to establish its test blueprint. This process ensures that our assessments always measure not only what is being taught in schools around the country, but also what demonstrably matters most for college and career readiness. No other assessment is built with the ability to continually assess what matters most, based on the most up-to-date evidence.

ACT's assessments provide the essential information to help get and keep students on the path toward readiness in the most efficient manner possible. Students in schools that administer ACT assessments, including the ACT and ACT Aspire, spend fewer than four hours taking the assessments, compared to as many as 7.5 hours for those taking other tests. In an era where over-testing is a significant concern, that's an important distinction.

The science behind ACT assessments—the evidence base and ongoing research—is critical to answering the key question of what matters most in 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 (The Condition of College & Career Readiness, Enrollment Management Trends Report, The Reality of College Readiness, and The Condition of STEM); and
- develop assessments, reports, and interventions that will help individuals navigate their personal path to success along a kindergarten-through-career continuum.

Accordingly, the following principles have shaped and will continue to drive ACT's development agenda:

1. Maximize instructional time.
2. Report results in instructionally relevant ways that support clear interpretation within and across content areas.
3. Establish reasonable testing times by assessing what research and evidence show to be the most critical factors for success after high school.
4. Leverage technology to enhance student engagement, produce more meaningful results, and share results in a timely fashion.
5. Increase the emphasis on evidence-centered design, implementing best practices as they mature, and improve ACT's capabilities within the highest-quality design and development processes.
6. Include science as a core academic domain in ACT's assessment batteries.
7. Reflect the reality that there are multiple dimensions of readiness and success (validated by research).

As a nonprofit educational research organization, ACT will use 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.

Survey Sample and Process

For the 2016 ACT National Curriculum Survey, ACT made online survey instruments available 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/writing, mathematics, reading (including English language arts and social studies), and science (including biology, chemistry, physics, and earth/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 numbers of survey respondents in each area.

Table 2.1 ACT National Curriculum Survey 2016 Respondents

Area	Number of Respondents
Early Elementary School	1,076
Late Elementary School	1,222
Middle School	1,331
High School	2,717
College	2,252
Supervisors	371
Employees	297
TOTAL	9,266

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 each content or skill in a given class they teach, while college instructors were asked to rate the importance of each content or skill as a prerequisite to success in a given class they teach.) ACT also asked the K–12 teachers to indicate whether or not they teach a particular content or skill and, if so, whether they teach it as a standard part of their course or as part of a review of material that should have been learned earlier. Some education participants were also asked other content-related questions depending on the grade level they teach.

Workforce participants were asked to rate discrete skills with respect to how important each is to entry-level success in the workplace. 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 a number of questions relevant to current education policy issues (e.g., assessments; technology; standards; student characteristics; and obstacles to success). These results are discussed in the companion report *ACT National Curriculum Survey 2016: Education and Work in a Time of Change* (www.act.org/research).

Because some content areas were surveyed in larger numbers than others, the values displayed in educational-level totals were averaged across English language arts, mathematics, and science. This ensured that, in these results, no one content area would have more influence than another.

2.4 Test Development Procedures

2.4.1 Review of Test Specifications

Two types of test specifications are used in developing the ACT tests: content specifications and statistical specifications.

Content specifications. Content specifications for the ACT tests were developed through the curricular analysis discussed above. While care is taken to ensure that the basic structure of the ACT tests remains the same from year to year, the specific characteristics of the test items used in each specification category are reviewed regularly. Consultant panels are convened to review both the tryout versions and the new forms of each test to verify their content accuracy and the match of the content of the tests to the content specifications. At these panels, the characteristics of the items that fulfill the content specifications are also reviewed. While the general content of the test remains constant, the particular kinds of items in a specification category may change slightly. The basic structure of the content specifications for each of the ACT tests is provided in Chapter 3.

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

The tests are constructed with a target mean item difficulty of about .58 for the ACT population and a range of difficulties from about .20 to .89. The distribution of item difficulties was selected so that the tests will effectively differentiate among students who vary widely in their level of achievement.

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 performance on the relevant mathematics test (i.e., the reporting category score, which will be introduced in Chapter 7).

2.4.2 Selection of Item Writers

Each year, ACT contracts with item writers to construct items for the ACT. The item writers are content specialists in the disciplines measured by the ACT tests and consist of ACT staff and outside contractors. 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.

Before being asked to write items for the ACT tests, potential item writer contractors (individuals and groups) are required to submit a sample set of materials for review. Each item writer receives an item writer's guide that is specific to the content area. The guides include examples of items and provide item writers with the test specifications and ACT's requirements for content and style. Included are specifications for fair portrayal of all groups of individuals, which includes avoidance of subject matter that may be unfamiliar to members of certain groups within society, a balanced representation for race/ethnicity, and gender-neutral language.

ACT Test Development staff evaluates each sample set submitted by a potential item writer. A decision concerning whether to contract with the item writer is made on the basis of that evaluation.

Each item writer under contract is given an assignment to produce a small number of items in the content area they are qualified for. The small size of the assignment ensures production of a diversity of material and maintenance of the security of the testing program, since any item writer will know only a small proportion of the items produced. Item writers work closely with ACT content specialists, who assist them in producing items of high quality that meet the test specifications.

Item-Writing Assignments

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 for the knowledge, skills, and abilities measured in each test. A typical assignment includes the evidenced-based item template, and focuses on a skill statement that the item needs to assess. Included in each template is a given a set of evidence statements that the item(s) must elicit.

Assignments are made available to qualified item writers 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. The information in the system can be connected to the template through the assignment.

2.4.3 Item Construction

The item writers must create items that are educationally important and psychometrically sound. A large number of items must be constructed because, even with good writers, many items fail to meet ACT's standards.

Each item writer submits a set of items, called a *unit*, in a given content area. Most mathematics test items are discrete (not passage based); some items may belong to a set of several items (e.g., several items based on the same paragraph or chart). All items on the English and reading tests are related to prose passages. All items on the science test are related to passages and/or other stimulus material, such as graphs and tables.

2.4.4 Review of Items

Content Review

After a unit is accepted, the unit is reviewed several times by ACT staff 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 teachers from each grade level, to participate in refining questions and verifying they are sampling constructs accordingly. 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 grades the items are intended to assess. During the external content review, items are evaluated for content accuracy, word count, item classification, item format, and language.

Bias, Sensitivity, Fairness, Accessibility Reviews

In order to verify that all items delivered to students are fair, unbiased, accessible, and non-offensive to all students, we conduct external fairness reviews for all items/tasks prior to pretesting and for forms before they become operational.

The external fairness review panel consists of experts in diverse educational areas who represent both genders and a variety of racial and ethnic backgrounds. Educators from appropriate grade levels and content areas participate and actively give us feedback. The fairness panel reviews items to help verify fairness to all students and to ensure that all items are free of bias or insensitivity. All comments are reviewed and appropriate changes are made. We select reviewers in a manner that no one state is over-represented because our stakeholders count on national representation to maintain the comparability of test forms and scores.

2.4.5 Item Tryouts

Items and passages that are judged to be acceptable in the review process are assembled into tryout units (sets of passages and items). These tryout units are then administered to different samples of the national examinee population. The samples of examinees are carefully selected to be representative of the total examinee population. Each sample of examinees is administered a tryout unit from one of the four academic areas covered by the ACT tests during an operational administration of the ACT, with the exception of the writing test which is generally pretested in a separate standalone tryout. The time limits for the tryout units permit the majority of students to respond to all items.

ACT pretests every item before it appears on an operational form to verify that the item is functioning properly.

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 percentile points in the distribution of those scores. These percentile points 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 well as the proportion in each group selecting each of the incorrect options. Biserial and point-biserial correlation coefficients of each tryout item are also computed.

Item analyses serve to identify statistically effective test items. Items that are either too difficult or too easy, and items that fail to discriminate between students of high and low educational achievement as measured by their corresponding ACT test scores, are eliminated or revised for future item tryouts. 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.

Each item is reviewed following the item analysis. ACT staff members scrutinize items flagged for statistical reasons to identify possible problems. Some items are revised and placed in new tryout units following further review. The review process also provides feedback that helps to improve the quality of items in the future.

2.4.6 Assembly of New Forms

Items that are judged acceptable in the review process are placed in an item pool. Preliminary forms of the ACT tests are constructed by selecting from this pool of items that match the content and statistical specifications for the tests.

For each test in the battery, items are selected to comply with the statistical specifications described in Chapter 3. The distributions of item difficulty levels obtained on recent forms of the four tests are displayed in Table 2.2. The data in Table 2.2 are taken from random samples of approximately 2,000 students from each of the six national test dates during the 2015–2016 academic year. In addition to the item difficulty distributions, item discrimination indices in the form of observed mean biserial correlations and completion rates are reported.

The completion rate is an indication of how speeded a test is for a group of students. A test is considered to be speeded if most 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 the six national test dates during the 2015–2016 academic year. The completion rate for each test is computed as the average proportion of examinees who answered each of the last five items.

Table 2.2 Difficulty[a] Distributions and Mean Discrimination[b] Indices for ACT Test Items, 2015–2016

	Observed difficulty distributions (frequencies)			
Difficulty range	English	Mathematics	Reading	Science
.00–.09	0	0	0	0
.10–.19	3	11	0	0
.20–.29	11	38	3	9
.30–.39	25	42	11	26
.40–.49	47	55	45	36
.50–.59	77	53	52	51
.60–.69	116	62	58	55
.70–.79	108	68	56	42
.80–.89	50	30	15	17
.90–1.00	13	1	0	4
No. of items[c]	450	360	240	240
Mean difficulty	0.63	0.55	0.61	0.58
Mean discrimination	0.56	0.59	0.54	0.53
Average completion rate[d]	92	93	94	95

[a]Difficulty is the proportion of examinees correctly answering the item.
[b]Discrimination is the item–total score biserial correlation coefficient.
[c]Six forms consist of the following number of items per test:
75 for English, 60 for mathematics, 40 for reading, and 40 for science.
[d]Completion rate is the proportion of examinees who answered each of the last five items.

2.4.7 Content and Fairness Review of Test Forms

The preliminary versions of the test forms are subjected to several reviews to ensure that the items are accurate and that the overall test forms are fair and conform to good test construction practice. 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 item correctly even though they lack knowledge in the subject areas or the required 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 level the form is intended to assess. These reviews follow a similar process to the item development external content review. Instead of the focus of the review being on individual items, however, the reviewers consider the quality of the form as a whole. They judge the form's content and cognitive distribution to make sure that there is no over- or under-representation in any category. Reviewers also look for the presence of "cluing" between items and other issues that could lessen the usefulness of the resulting scores.

Additionally, all newly-developed ACT forms also must go through an external content and fairness panel review. This panel consists of experts in diverse areas of education with a balanced representation of genders and have experience working with diverse populations. The fairness panel reviews the forms to help ensure that all items are free of bias or insensitivity for all examinees.

After the panels complete their reviews, ACT summarizes the results. All comments from the consultants are reviewed by ACT staff members, 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 printing.

In all, at least 16 independent reviews are made of each test item before it appears on an operational form of the ACT. The many reviews are performed to help ensure that each student's level of achievement is accurately and fairly evaluated.

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 or minimize the influence of the problem item as necessary. 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 procedures are conducted on the test data. DIF can be described as a statistical difference between the probability of the specific population group (the "focal" group) answering the item correctly and the comparison population group (the "reference" group) getting the item right given that both groups have the same level of achievement with respect to the content being tested. The procedures currently used for the analysis include the standardized difference in proportion-correct (STD) procedure and the Mantel-Haenszel common odds-ratio (MH) procedure.

Both the STD and MH techniques are designed for use with multiple-choice items, and both require data from significant numbers of examinees to provide reliable results. In the analysis of items in an ACT form, large samples representing examinee groups of interest (e.g., males and females) are selected from the total number of examinees taking the test. The examinees' responses to each item on the test are analyzed using the STD and MH procedures. Compared with pre-established criteria, the items with STD or MH values exceeding the tolerance level are flagged. The flagged items are then reviewed further by the content specialists for possible explanations of the unusual STD or MH results. In the event that a problem is found with an item, actions will be taken as necessary to eliminate or minimize the influence of the problem item.

2.5 Test Development Procedures for the Writing Test

This section describes the procedures that are used in developing essay prompts for the ACT writing test. These include many of the same stages as those used to develop the multiple-choice tests.

2.5.1 Prompt Writers

Each year, ACT writing specialists generate prompt ideas and develop the resultant prompts. ACT writing specialists have deep 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 that offer adequate complexity and depth so that examinees can write a thoughtful and engaging essay. Topics are carefully chosen so that they are neither too vast nor too simplistic so they do not require specialized prior knowledge. The topics are designed so that a student should be able to respond to a topic 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 then refined by ACT writing specialists, the prompts go through a rigorous review process by 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 for 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 each year to evaluate new potential ACT writing prompts to select those suitable for operational use. Students from rural and urban settings, small and large schools, and both public and private schools write responses to the new prompts, which are then read and scored by trained ACT readers.

Prompts are evaluated from both content and statistical perspectives to ensure scores 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 the 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 analyzed for acceptability, validity, and accessibility. The new field-tested prompts are also reviewed to ensure that they are compatible with previous operational prompts, that they function in the same way as previous prompts, and that they adhere to ACT's rigorous standards.

To ensure the comparability of the 2–12 overall writing scores, prompts are selected for operational use if they perform similarly to the anchor prompts, meaning the distributions of 2–12 scores are similar across the prompts. A similar procedure had been used to ensure the comparability of the ACT writing scores prior to fall 2015.

2.6 ACT Scoring Procedures

For each of the content tests in the ACT (English, mathematics, reading, and science), the raw scores (number of correct 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. In addition to the four ACT test scores and Composite score, several reporting category scores are reported. ACT reporting categories are aligned with 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. The number of items for a particular reporting category can vary across different test forms. Because these scores are raw scores, they are not directly comparable across different test forms. For each reporting category, the score report shows the total number of points possible, the number of correct responses, the percent of correct responses, and the ACT Readiness Range. The ACT Readiness Ranges enable students to see at a glance how their performance on each reporting category compares to students who have met the ACT College and Career Readiness Benchmark for that specific subject. For each reporting category, the ACT Readiness Range was calculated by regressing the percentage of points achieved on students' scale scores for the corresponding subject. The minimum value of the range is the point that corresponds to the predicted percentage of points that would be achieved by a student who meets the ACT College and Career Readiness Benchmark on the overall subject test. The maximum value of the range corresponds to answering all questions in that reporting category correctly. 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 to more clearly show which areas require the most assistance for additional learning and intervention.

In addition to the above scores, if the student took the writing test, the student's essay is read and scored independently by two trained readers using a six-point scoring rubric with four domains. Essays are evaluated on the evidence they demonstrate of student ability to clearly state one's own perspective on the issue and analyze the relationship between that perspective and at least one other perspective, develop and support ideas with reasoning and examples, organize ideas clearly and logically, and communicate ideas effectively in standard written English. Essays are scored analytically—that is, on the basis 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 2–12, is the rounded average of the four domain scores. A student who takes the writing test also receives an English Language Arts (ELA) score on a score scale ranging from 1 to 36. Writing test results do not affect a student's Composite score.

2.6.1 Additional Scores and Indicators

In September 2015, ACT began reporting Science, Technology, Engineering, and Mathematics (STEM) scores, a combination of students' mathematics and science scores, and English Language Arts (ELA) scores, a combination of students' English, reading, and writing scores for students who take the writing test, in addition to the four multiple-choice test scale scores and Composite scores. Also introduced at that time were Understanding Complex Texts (UCT) and Progress Toward the ACT® WorkKeys® National Career Readiness Certificate® (NCRC®) indicators.

2.6.2 STEM and ELA Scores

The STEM score is the average of the 1–36 mathematics and science scale scores rounded to the nearest integer (fractions of 0.5 or greater round up). Only students who receive scores for both tests receive a STEM score.

The combined ELA score is the rounded average of the English, reading, and the 1–36 writing scale scores. Only students who take the English, reading, and writing 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. Procedures for obtaining the 1–36 writing scale scores are described in Chapter 9.

2.6.3 Understanding Complex Texts Indicator

ACT test score reports include a UCT 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 demand. This indicator is based on scores from a subset of items on the reading test. These items measure students' global comprehension of the passages instead of sentence- or word-level understanding. Student performance on these items is classified into three performance levels: Below Proficient, Proficient, and Above Proficient.

2.6.4 Progress Toward the ACT National Career Readiness Certificate Indicator

The Progress Toward the ACT NCRC indicator is based on students' ACT Composite scores. It provides an estimate of students' likely performance on the WorkKeys NCRC. The WorkKeys NCRC is an assessment-based credential that certifies foundational work skills important for job success across industries and occupations. The ACT NCRC is based on the results of three ACT WorkKeys assessments: applied mathematics, graphic literacy, and workplace documents. Scores on these assessments determine the certificate level—no certificate, Bronze, Silver, Gold, or Platinum—an individual can earn. 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 http://workforce.act.org/credential. More details on the ACT test scores and indicators can be found in Chapter 7.

2.6.5 Scoring Appeals and Inquiries

Electronic scanning devices are used to score the four multiple-choice tests of the ACT, thus minimizing the potential for scoring errors. If a student believes that a scoring error has been made, ACT hand-scores the answer document (for a fee) upon receipt of a written request from the student. Strict confidentiality of each student's record is maintained.

If a student believes that a writing test essay has been incorrectly scored, that score may be appealed, and the essay will be reviewed and rescored (for a fee) by two new expert readers. The two new readers score the appealed essay without knowledge of the original score, and the new score is adjudicated by ACT staff writing specialists before being finalized.

For certain test dates (specified in the current year's booklet *Registering for the ACT,* also found online at www.act.org), examinees may obtain (upon payment of an additional fee) a copy of the test items used in determining their scores, 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 (surrogate testing), unusual similarities in answers of examinees at the same test center, or other indicators that the test scores may not accurately reflect the examinee's level of educational achievement—including but not limited to examinee misconduct—may result in ACT's canceling the test scores. For a detailed description of how ACT handles score cancellations, refer to ACT's Terms and Conditions of Registration (http://www.act.org/the-act/terms).

References

ACT. (2007, 2009, 2013, 2016). *ACT National Curriculum Survey® 2016.* Iowa City, IA: Author.

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: American Council on Education, and Macmillan Publishing Company.

Spray, J. A. (1989). Performance of three conditional DIF statistics in detecting differential item functioning on simulated tests (ACT Research Report No. 89-7). Iowa City, IA: ACT, Inc.

Chapter 3

Test Specifications

3.1 Overview

The ACT test is constructed to meet specifications for content balance within the assessment domain. Test specifications define ranges of the number of items in different content categories and at levels of cognitive complexity. These content blueprints ensure that the knowledge and skills in the content domains are sampled consistently across parallel forms of the tests. 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 asks students to assume the role of a writer who analyzes texts and makes decisions to revise and edit the writing. The test measures understanding of the conventions of standard written English (punctuation, grammar and usage, and sentence structure), production of writing (topic development, organization, unity, and cohesion), and knowledge of language (word choice, style, and tone). The test consists of five essays, or 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 passage 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, vocabulary, and rote recall of rules of grammar are not tested.

Some items refer to underlined or highlighted portions of the passage and offer several alternatives to the designated portion. These items include “NO CHANGE” to the designated portion in the passage as one of the possible responses. Some items are identified by a number or numbers in a box. These items ask about a section of the passage or about the passage as a whole. The student must decide which choice best answers the item posed.

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 across the nation in many educational contexts. The ACT English test assesses skills across a range of cognitive complexity using items at DOK Levels 1, 2, and 3. All multiple-choice items are classified by ACT content experts according to the following level descriptions.

Table 3.1 Level Descriptions for English

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.

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 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 and the approximate percentage of the test items in each reporting category are given below. In addition, the overall English test score, along with the reading test score and the writing scale score, is used to determine the ELA score (see Chapter 7 for more description in forming the ELA score).

Production of Writing

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

- **Topic Development**
  Students demonstrate an understanding of, and control over, the rhetorical aspects of texts by identifying the purposes of parts of texts, determining whether a text or part of a text has met its intended goal, 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 through ensuring precision and concision in word choice and maintaining consistency in style and tone.

Conventions of Standard English

Students apply an 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 relationships between and among clauses, placement of modifiers, and shifts in sentence construction.
- **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 Specification Ranges by Reporting Category for English

	Number of Items	Percentage of Test
Reporting Category		
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 considers the whole of a student's mathematical development up through topics typically taught at the beginning of Grade 12 in US schools, focusing on prerequisite knowledge and skills important for success in college mathematics courses and career training programs. The domain is divided into recent topics 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 fluency is required; most students have sufficient time to complete the test. 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, and so that all problems can be done without a calculator in a reasonable amount of time.

Students have 60 minutes to complete 60 multiple-choice items. Each item has five response options, and students are instructed to choose the correct option. The test contains problems ranging from easy to very challenging in order to reliably report on readiness levels for students with different preparation. Extended accessibility supports provide for fair and comparable mathematics scores across a range of circumstances. More information on accessibility can be found in Chapter 4.

Cognitive Complexity and Depth of Knowledge (DOK)

Being judgment-based, individual DOK coding of items tends to differ from group to group and time to time; therefore ACT incorporates substantial training, discussion, and multiple inputs to achieve a consistent implementation of DOK. Development targets in terms of DOK provide parallelism from test form to test form as well as ensuring a mix of cognitive complexity essential for measuring mathematics achievement. Because this is a rough-grained measure, ACT staff look at a finer-grained level when constructing forms to ensure good diversity across tasks.

Table 3.3 DOK Level Descriptions for Mathematics

Depth of Knowledge Level	Description
DOK1	Recall and reproduction
DOK2	Skills and concepts
DOK3	Strategic thinking

3.3.2 Mathematics Scores and Reporting Categories

The mathematics test score is based on all 60 items. This score is reported on the ACT mathematics scale, which ranges from 1 to 36. Properties of this scale are given in Chapter 7.

The mathematics test score provides a powerful interpretation based on the successes of similar students over past decades. A comparison to the ACT College Readiness Benchmark for mathematics (currently a score of 22) gives a general idea about success in a typical postsecondary College Algebra course. (Individual colleges have tailored interpretations of the mathematics test score in terms of placement and course success for a number of their courses.) The ACT College and Career Readiness Standards show combinations of mathematical skills likely for students with a given mathematics test score. Normative information allows interpretation relative to classmates, students in the same state, and a standard population of ACT test takers. More information is available in Chapter 8 and on the ACT website at www.act.org.

The STEM score is the average of the mathematics test score and the science test score (only available for students who get scores on both tests). The STEM score is related to success in postsecondary science, technology, engineering, and mathematics courses. More information about the STEM score can be found in Chapter 7.

There are eight additional reporting categories, designed to give more detail about a student's mathematical achievement. A student's mathematics test score corresponds to information about the group of all students with that score; 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.

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**
  Coming into high school, students have some knowledge of the real number system. Because they have an understanding of and fluency with rational numbers and the four basic operations, they can work with irrational numbers by manipulating rational numbers that are close. Students are ready to move from integer exponents to rational exponents and are also ready to probe deeper into properties of the real number system. Students extend their knowledge to include complex numbers, which offer the solutions to some simple equations that have no real-number solutions, and students learn to compute in this system. Students go further, exploring properties of complex numbers—and in the process learn more about real numbers. Students explore vectors and matrices and view them as number systems with properties, operations, and applications.

- **Algebra**
  Students coming into high school build on 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. They continue to make sense of expressions in terms of their parts in order to use their fluency strategically and to solve problems. Through repeated reasoning, students develop a general understanding of solving equations as a process that provides justification that all the solutions will be found. Students extend their proficiency to equations such as quadratic, polynomial, rational, radical, and systems, integrating an understanding of solutions in terms of graphs. Families of equations have properties that make them useful for modeling. Polynomials form a system analogous to adding, subtracting, and multiplying integers; solutions of polynomial equations are related to factors of a polynomial. Students recognize these relationships in applications and create expressions, equations, and inequalities to represent problems and constraints. Students see rational expressions as a system analogous to rational numbers, apply the binomial theorem, and solve simple matrix equations that represent systems of linear equations.

- **Functions**
  Functions have been with students since their early years: consider the counting function that takes an input of "seven" and gives "eight" and an input of "twelve" to give "thirteen." Understanding general properties of functions will equip students for problem solving with new functions they create over their continued studies and careers. Functions provide a framework for modeling real-world phenomena, and students become adept at interpreting the characteristics of functions in the context of a problem and become attuned to differences between a model and reality. Some functions accept all numbers as inputs, but many accept only some numbers. Function notation gives another way to express functions that highlights properties and behaviors. Students work with functions that have no equation, functions that follow the pattern of an equation, and functions based on sequences, which can even be recursive. Students investigate particular families of functions—like linear, quadratic, and exponential—in terms of the general function framework: looking at rates of change, algebraic properties, and connections to graphs and tables, and applying these functions in modeling situations. Students also examine a range of functions like those defined in terms of square roots, cube roots, polynomials, exponentials, logarithms, and trigonometric relationships, and also piecewise-defined functions.

  Students see solving an equation in terms of an inverse function. Students have seen shifts in graphs due to parameter changes, but now they develop a unified understanding of translations and scaling through forms such as $f(x - c)$, $f(x) + c$, $a\,f(x)$ and $f(-ax)$. Students connect the trigonometry of right triangles to the unit circle to make trigonometric functions, and they explore algebraic relationships among these functions. They use these functions to model periodic behavior.

  Students graph rational functions and learn about asymptotes. They compose functions in other ways besides translation and scaling, going deeper into how inverse functions apply to solving equations with more than one solution, in particular for trigonometric functions. They explore algebraic properties of trigonometric functions such as angle addition properties.

- **Geometry**
  Starting from an understanding of congruence and rigid motions, students add depth to what they know about dilations and add precision to their understanding of similarity. Students make constructions, solve problems, and model with geometric objects. Informal arguments give a chain of reasoning that leads to formulas for the area of a circle and then on to volume of cylinders, pyramids, and cones. Through the lens of similar triangles, students understand trigonometric ratios as functions of the angle and they solve right-triangle problems. All these results transfer to the coordinate plane, where analytic treatment of distance allows students to derive conditions for parallel and perpendicular lines, to split a line segment into pieces with a given ratio of lengths, to find areas, and to develop equations for circles and for parabolas that have a directrix parallel to an axis.

  Students go further into trigonometry, deriving a formula for the area of a general triangle in terms of side lengths and the sine of an angle, moving on to the law of sines and law of cosines, which give straightforward answers to items about nonright triangles. Students derive equations for ellipses and hyperbolas. Students use Cavalieri's principle to justify formulas, such as the formula for volume of a sphere.
- **Statistics & Probability**
  In high school, students learn about the role of randomness in sample surveys, experiments, and observational studies. Students use data to estimate population mean or proportion and make informal inferences based on their maturing judgment of likelihood. They can compare qualities of research reports based on data and can use simulation data to make estimates and inform judgment.

  Before high school, students have tacitly used independence, but now the idea is developed with a precise definition. Students relate the sample space to events defined in terms of "and," "or," and "not," and calculate probabilities, first using empirical results or independence assumptions, and later using the ideas of conditional probability. Students understand the multiplicative rule for conditional probability and study permutations and combinations as a tool for counting. Students 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 decisions.

Integrating Essential Skills

This reporting category focuses on whether students can put together understandings and skills to solve problems of moderate to high complexity. Topics include rate and percentage; proportional reasoning; area, surface area, and volume; average and median; quantities and units; expressing numbers in different ways; using expressions to represent quantities and equations to capture relationships; the basics of functions; congruence, symmetry, and rigid motions; data analysis and representation; associations between two variables; and model fit. In addition to learning more content, students should grow in sophistication, accumulating and applying skills in higher order contexts. Students should be able 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. In order to assess whether students have had appropriate growth, the items in this reporting category are at least DOK Level 2, with a significant portion at DOK Level 3. DOK is judged relative to well-prepared high school students in grade 11–12.

Modeling

Modeling uses mathematics to represent, through a model, an analysis of an actual, 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 using 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 of 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 always available on the ACT website at www.act.org.

3.3.4 Item Sets

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 2–5 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 for another item.

3.3.5 Mathematical Practices

Mathematical practices highlight crosscutting mathematical skills and understandings and the complex and vital ways the skills and understandings integrate with content. Test items focus on important mathematics, which includes various levels of expertise with mathematical practices. Therefore, scores include mathematical practices. The Modeling score pulls out that particular mathematical practice across a variety of contexts.

3.3.6 Mathematics Test Blueprints

Table 3.4 below summarizes content constraints for the mathematics test. Test construction also takes into account coverage and variety within each of the categories. Each form is built to have a similar distribution of item percentage-correct values, based on predictions made from pretest performance. Pretest item discrimination statistics must be sufficiently high. Form balance is examined in a number of areas such as word count, and substitutions are made as appropriate.

PHM and IES are specified separately in order to capture the spirit of those categories. As explained above, PHM represents the newer topics, and the assessment includes the whole range DOK1–DOK3. IES represents topics that should be very familiar and what is important for college readiness measures is putting these familiar skills to work in higher complexity tasks.

Table 3.4 Specification Ranges by Reporting Category for Mathematics

	Number of Items	Percentage of Test
Reporting Category		
Preparing for Higher Mathematics	34–36	57–60%
• Number & Quantity	4–6	7–10%
• 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	≥ 16	≥ 27%
Total Mathematics	60	100%

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. The passages in the reading test include both literary narratives and informational texts from the humanities, natural sciences, and social sciences. Passages are representative of the kinds of text commonly encountered in first-year college curricula. Each passage is preceded by a heading that identifies what type of passage it is (e.g., “Literary Narrative”), 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 multiple-choice test items. The test 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. 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 across a range of cognitive complexity using items at DOK Levels 1, 2, and 3. All multiple-choice items are classified by ACT content experts according to the following level descriptions.

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.

3.4.2 Reading Scores and Reporting Categories

Five scores are reported for the ACT reading test: a total test score based on all 40 items, three reporting category scores based on specific knowledge and skills, and an Understanding Complex Texts indicator. The three reporting categories addressed in the reading test are Key Ideas & Details; Craft & Structure; and Integration of Knowledge & Ideas. In addition, the overall reading test score, along with the English test score and the writing scale score is used to determine the ELA score (see Chapter 7 for more description in forming the ELA score). A description and the approximate percentage of the test devoted to each reporting category are given below.

Key Ideas & Details

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

Craft & Structure

Students determine word and phrase meanings, analyze an author's word choice rhetorically, analyze text structure, understand authorial purpose and perspective, and analyze characters' points of view. They interpret authorial decisions rhetorically 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 that are related by topic. Some items will require students to analyze how authors construct arguments, evaluating reasoning and evidence from various sources.

3.4.3 Reading Test Blueprints

Table 3.6 Specification Ranges by Reporting Category for Reading

	Number of Items	Percentage of Test
Reporting Category		
Key Ideas & Details	22–24	55–60%
Craft & Structure	10–12	25–30%
Integration of Knowledge & Ideas	6–8	15–18%
Total Number of Items	**40**	**100%**

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 Science/Space Science.

Students are assumed to have a minimum of two years of introductory science, which ACT's National Curriculum Survey has identified as typically one year of Biology and one year of Physical Science and/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 different 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 addressing the same scientific phenomenon that are inconsistent with one another).

The Nature of the ACT Science Test: What does it measure?

The ACT science 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.

These three domains, and the knowledge and skills encompassed in each domain, were derived from ACT's decades of empirical data and research on college and career readiness in science. The domains and their skills comprise the ACT College and Career Readiness Standards for science, which link specific skills and knowledge with quantitatively determined score ranges for the ACT science test and the Benchmark in science that is predictive of success in science at the postsecondary level. These three domains are also the reporting categories for the ACT science test (see Table 3.8). ACT also reviews Benchmarks in science and standards from state standards documents as well as national (e.g., the Next Generation Science Standards) and international standards documents and monitors the impact of these documents on science curricula to assure alignment and, when needed, to evolve the constructs of the test. All items on the ACT science test are based on authentic scientific scenarios that are built around important scientific concepts and are designed to mirror the experiences of students and working scientists engaging in real science. The ACT science test focuses on multidimensional assessment (to measure 3-dimensional learning in science), with items that require students to apply multiple domains. Some of the items require that the students have discipline-specific content (e.g., knowledge specific to an introductory high school physical science or biology course), but all of the items focus on science process skills. Research conducted by ACT on science curricula and instruction at the high school and postsecondary levels shows that while having fundamental understanding of disciplinary science content knowledge is important, being able to apply science practices/process skills to science content to solve problems is more strongly tied to college and career readiness in science. The ACT science test focuses on measuring the science skills and knowledge that are empirically tied to college and career readiness.

Cognitive Complexity and Depth of Knowledge

The ACT science test assesses 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 have quite captured the different dimensions associated with items focused on science skills and practices. Below is an example of how items on the ACT science test are classified by DOK.

Table 3.7 DOK Level Descriptions for Science

Depth of Knowledge Level	Description
DOK1	Locating/reproducing information
DOK2	Applying skills and concepts
DOK3	Integrating skills and concepts (strategic thinking)

3.5.2 Science Scores and Reporting Categories

Four scores are reported for the ACT science test, including a science test score based on all 40 items, and three reporting category scores based on different domains of scientific knowledge, skills, and practices. 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, and the percentage of the test devoted to each reporting category is provided in Table 3.8. The overall score on the science test is also used, with the mathematics score, to determine the STEM score.

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 Specification Ranges by Reporting Category for Science

	Number of Items	Percentage of Test
Reporting Category		
Interpretation of Data	18–22	45–55%
Scientific Investigation	8–12	20–30%
Evaluation of Models, Inferences & Experimental Results	10–14	25–35%
Passage Formats		
Data Representation	12–15	30–40%
Research Summaries	18–21	45–55%
Conflicting Viewpoints	6–8	15–20%
Total Number of Items	40	100%

Table 3.9: Specification Ranges by Science Content Area Specifications

Content Area	Number of Passages	Number of Items	Percentage of Test
Biology	2	12–14	30–35%
Chemistry	1–2	6–14	15–35%
Physics	1–2	6–14	15–35%
Earth/Space Science	1–2	6–14	15–35%
Total	6	40	100%

3.6 Writing Test

3.6.1 Description of the Writing Test

The ACT writing test is a 40-minute essay test that measures students' writing skills—specifically those skills emphasized in high school English classes and in entry-level college composition courses. The information from the writing test tell postsecondary institutions about students' ability to think critically about an issue, consider different perspectives on it, and compose an effective argumentative essay in a timed condition. An image of the essay will be available to the student's high school and the colleges selected for score reporting.

The writing test underwent a number of enhancements that became operational in September 2015. The enhanced 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 DOK Level Descriptions for Writing

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.

3.6.2 Writing Scores and Domains

Students who take the optional writing test receive a total of five scores: a single subject-level writing score reported on a range of 2–12[1] and four domain scores, also on a range of 2–12, that are based on an analytic scoring rubric. The subject-level score is the rounded average of the four domain scores.

Taking the writing test does not affect the student's subject area scores or Composite score. However, a writing test score, along with the overall English and reading test scores are needed to report an ELA score.

[1] Students who took the writing test from September 2015 to June 2016 received a subject-level writing score reported on a 1–36 scale and not the 2–12 subject-level score that is currently reported. It should also be noted that the current 2–12 subject-level writing score is not the same as the 2–12 score 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. The current test is also 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.

The four domain scores on the writing test are Ideas & Analysis, Development & Support, Organization, and Language Use & Conventions. Two trained raters score each essay on a scale of 1–6 in each of the four domains. Each domain score represents the sum of the two raters' scores using the analytic rubric in Table 3.12. If the ratings disagree by more than one point, a third rater evaluates the essay and resolves the discrepancy.

Ideas & Analysis

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

Development & Support

Scores in this domain reflect the ability to discuss ideas, offer rationale, and strengthen an argument. Competent 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 purposes. Organizational choices are integral to effective writing. Competent writers arrange their essay in a way that clearly shows the relationship among ideas, and they guide the rater through their discussion.

Language Use & Conventions

Scores in this domain reflect the ability to use written language to clearly convey ideas. Competent 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.

3.6.3 Performance Scoring

Various performance scoring processes and procedures are utilized for scoring the ACT Writing test, such as rangefinding, rater training and qualification, as well as 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 in order to continue to 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 rater accuracy, 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 rangefinding, training development, and ongoing calibration.

3.6.3.1 Rater Training and Qualification

The rangefinding process is the basis of scoring criteria validation and the development of effective rater training materials. A panel of assessment and content experts meets to review a sample of student responses and ensures 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.

Development of 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. In addition to anchor sets, ACT’s rangefinding panels also develop practice and qualifying sets.

Rater candidates are introduced to the rubric and the writing prompt, and then 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 in order to identify appropriate scores. Annotated feedback is provided at the conclusion of each practice set.

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

Rater training and qualification use a selected “baseline” prompt. Baseline training with qualification is administered at least twice annually for all raters. After qualifying, additional writing prompts are introduced 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.

3.6.3.2 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 a number of quality assurance processes that establish and maintain a 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:

- Reliability scoring: Every ACT Writing response is reviewed and scored by at least two independent, qualified raters. In cases where scores are non-adjacent, a response is automatically rerouted for a third review read by a Scoring Supervisor or Director and the discrepancy is appropriately resolved. Due to the rigorous training and qualification requirements, non-adjacency rates routinely amount to less than 4% of the overall response population.
- Validity: Validity responses are selected and pre-scored by Scoring Supervisors and Scoring Directors, and 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 Supervisor feedback and calibration tests. A rater who fails to demonstrate improved accuracy may be released from the project and his or her work reset and rescored.
- Backreading: The backreading process enables Supervisors and Directors to review raters' work and provide effective, tailored feedback based on specific scoring examples. The backreading process also allows for the application of new scores 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 in order to maintain rater accuracy and to address any emergent scoring trends. Calibration sets are compiled by Supervisors and Directors to address specific scoring trends, or as 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 3.11 provides a sample of some of the available reports.

Table 3.11 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.

Table 3.12 Writing Test Analytic Scoring Rubric

	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.

Table 3.12 Writing Test Analytic Scoring Rubric—continued

	Ideas & Analysis	Development & Support	Organization	Language Use & Conventions
Score 4: **Responses at this score point demonstrate adequate skill in writing an argumentative essay.**	The writer generates an argument that engages with multiple perspectives on the given issue. The argument's thesis reflects clarity in thought and purpose. The argument establishes and employs a relevant context for analysis of the issue and its perspectives. The analysis recognizes implications, complexities and tensions, and/or underlying values and assumptions.	Development of ideas and support for claims clarify meaning and purpose. Lines of clear reasoning and illustration adequately convey the significance of the argument. Qualifications and complications extend ideas and analysis.	The response exhibits a clear organizational strategy. The overall shape of the response reflects an emergent controlling idea or purpose. Ideas are logically grouped and sequenced. Transitions between and within paragraphs clarify the relationships among ideas.	The use of language conveys the argument with clarity. Word choice is adequate and sometimes precise. Sentence structures are clear and demonstrate some variety. Stylistic and register choices, including voice and tone, are appropriate for the rhetorical purpose. While errors in grammar, usage, and mechanics are present, they rarely impede understanding.
Score 3: **Responses at this score point demonstrate some developing skill in writing an argumentative essay.**	The writer generates an argument that responds to multiple perspectives on the given issue. The argument's thesis reflects some clarity in thought and purpose. The 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.

Table 3.12 Writing Test Analytic Scoring Rubric—continued

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

References

Webb, N. L. (2002, March 28th). Depth-of-knowledge levels for four content areas. Unpublished manuscript. Retrieved from: http://facstaff.wcer.wisc.edu/normw/All%20content%20areas%20%20DOK%20levels%2032802.pdf

Chapter 4

Accessibility

4.1 Equal Opportunity for All

The Standards for Educational and Psychological Testing (AERA, APA, & NCME, 2014), address fairness in testing as the central concern posed by the threat to validity known as measurement bias. The Standards specify two major concepts that have emerged in the literature for minimizing such bias, namely accessibility and universal design. Accessibility is defined as "the notion that all test takers should have an unobstructed opportunity to demonstrate their standing on the construct(s) being measured" (p.49). The second major concept, universal design, is defined as "an approach to test design that seeks to maximize accessibility for all intended examinees" (p.50).

Accessibility is an inclusive concept that recognizes that the need for personalized communication supports is not restricted to any one group of examinees. It describes needs all individuals have, regardless of whether one is identified with a specific diagnostic label. Accessibility refers to the needs of the entire testing population, including those students identified as having disabilities and those who are English learners (ELs), as well as those who are not identified with any diagnostic label at all. All these groups have in common a shared need to be able to fairly and effectively communicate what they know and can do when they take a test. The ACT's system of accessibility encompasses this entire spectrum of user needs for communication tools and support.

The theory of action that ACT implements, known as Access by Design (Fedorchak, 2013), incorporates elements of Universal Design for Learning (UDL) described by the Center for Applied Special Technologies (CAST, 2011) and Evidence Centered Design (Mislevy, Almond, & Loves, 2004; Mislevy & Haertel, 2006) into its conceptual structure. With clearly defined construct(s) being measured in the assessments, ACT test design teams are able to collaborate with content teams, accessibility specialists, and researchers to design assessments in a way that supports the effective communication of all test takers and removes construct-irrelevant barriers to meet the needs of and provide a fair performance pathway for all learners, including populations with diverse needs.

The schematic representation in Figure 4.1 from Access by Design (Fedorchak, 2013) reflects the basic cognitive processing and communication that occurs during testing: stimulus presentation and receptive communication, internal processing, and expressive communication through response production. These critical communication access points are designed into the ACT's measurement architecture to achieve meaningful accessibility for all participating.

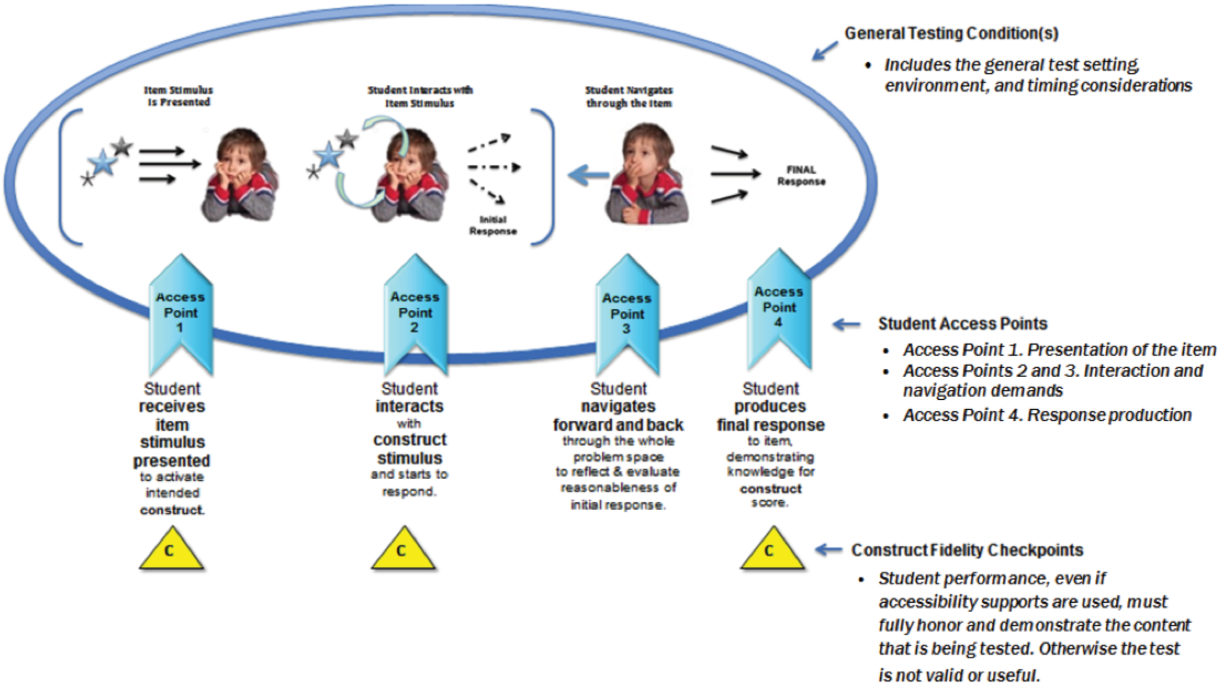


Figure 4.1 ACCESS POINTS: How an examinee experiences a test item.

Every examinee within the examinee population should have the opportunity to demonstrate his or her knowledge, skills, and abilities on the defined construct. When an examinee is unable to access the test tasks or items because construct-irrelevant barriers block his or her performance, the assessment becomes fundamentally unfair and the validity of the assessment threatened. The likelihood that an examinee will encounter a construct-irrelevant barrier to test performance increases for those with certain disabilities or for others who have limited communication pathways available, such as those who are just learning the English language. An important component of designing assessments therefore becomes developing and administering them in ways that remove accessibility barriers without compromising the validity of test score inferences (Hansen & Mislevy, 2008).

In the ACT high-level test design process, ACT first reviews its model of content to ask:

1. **What claim(s) does the performance on this item support?**
   Claim level 1: Content area, broad constructs
   Claim level 2: Broad subarea within content
   Claim level 3: Primary claim
   Claim level 4: Secondary claim, as applicable

Then, ACT reviews the proposed high-level test design and detailed item specifications to ask:

2. **What communication performance does this task require of the learner?**
   What are the presentation demands (for receptive communication)?
   What are the interaction and navigation demands (processing) demands?
   What are the response demands (for expressive communication)?
   What are the general test conditions required of the learner and will they support alternate allowable communication modes?

Finally, moving into the Detailed Level Design (DLD), given what is known about the test's content and design elements, ACT considers every learner population to be included in the ACT test takers and asks:

3. **Who has a valid and fair pathway to demonstrate the required performance?**
   Default learners? If no, what are the gaps?
   Learners with low or impaired vision? If no, what are the gaps?
   Learners with blindness? If no, what are the gaps?
   Learners who are deaf or hard of hearing? If no, what are the gaps?
   English learners? If no, what are the gaps?
   Learners with fine or gross motor impairment? If no, what are the gaps?
   Learners with reading, language, or mathematics notational impairment? If no, what are the gaps?
   Learners with attention, behavior, or other health impairment? If no, what are the gaps?

During this DLD process, test and item specifications are created that capture accessibility features and concerns. During this process, ACT defines the acceptable range of communication formats permitted during the test that will honor each of the constructs while allowing accessible communication and performance pathways for all intended users. For example, in what ways can writing be communicated and demonstrated? In what ways can independent calculation be demonstrated? What ways of communicating will NOT demonstrate reading? What ways of communicating will NOT demonstrate writing? This process of defining acceptable communication limits serves to further clarify the essential components of construct definition.

Another aspect of creating item specifications is determining which item interaction formats will most effectively measure the constructs defined while also allowing the widest, most diverse group of learner populations to engage with and respond to those items. There are many ways to ask a question or to pose a task (e.g., multiple choice, text entry, matching, graphing, and other interaction formats). Models of item interaction, once selected to measure the constructs, are then reviewed to determine which item interaction formats might pose barriers to communication and performance for some intended users. When item interaction barriers are identified, a review of alternate methods of communicating the construct and asking the question are considered. Where possible, instead of changing the interaction format, allowable communication supports that will remove unnecessary performance barriers are identified.

A culminating DLD activity is the defining of allowable accessibility features. Allowable accessibility features are those forms of communication support and tools that, if used, will fully honor the constructs defined, enabling the examinee to demonstrate the construct as it is designed to be measured. Such communication supports and tools (accessibility features) do not do anything for the examinee that the examinee should be doing for him- or herself independently. These allowable accessibility features simply remove construct-irrelevant barriers to performance of the construct. They level the playing field for all, creating a fair communication opportunity without giving one examinee an advantage over any other.

The DLD team comprises content specialists, accessibility specialists, and research and design specialists. The DLD team evaluates whether allowing use of the feature in appropriate situations enables measurement of the intended construct, or does allowing the use of the feature result in a distortion or violation of the defined construct being measured, which may lower validity and adversely impact fairness.

Implementing these principles of universal design is essential, but not sufficient by itself to meet the needs of every learner with diverse needs (CAST, UDL Guidelines, Version 2.0, 2011). To build an assessment system that meets the needs of all populations tested and provides a fair communication and performance pathway for all learners, more than one level of support is needed. The ACT has established a continuum of supports for effective communication that spans from the most simple, common accessibility tools used by everyone, to the most intensive accessibility supports that require the user to have specific qualifications and expertise.

Two levels of accessibility supports are currently permitted for the ACT: (1) embedded-universal tools and (2) allowed accommodations. Embedded-universal tools are commonly used by many people, available to all examinees, and do not need to be requested in advance. Allowed accommodation-level supports and tools are the most intensive level of support. Qualifying for use of intensive accommodation-level supports involves the official documentation of user need for the support, and a history of successful use of that support by the learner. Certain supports are only available with the paper format of the test and are outlined later in this chapter.

4.2 Test Administration and Accessibility Levels of Support

4.2.1 Understanding Levels of Accessibility Support

The ACT test has multiple levels of accessibility supports. These accessibility supports:

- allow all examinees to gain access to effective means of communication that in turn allow examinees to demonstrate what they know without providing an advantage over any other examinee
- enable effective and appropriate engagement, interaction, and communication of examinee knowledge and skills

- honor and measure academic content as the test developers originally intended
- remove unnecessary barriers to examinees' demonstrating the content, knowledge, and skills being measured on the ACT

In short, accessibility supports do nothing for the examinee academically that he or she should be doing independently; the supports just make interaction and communication possible and fair for each examinee.

4.2.2 Accessibility Support Structure

The ACT's accessibility system structure defines three potential levels of support that range from minor support (embedded-universal system tools) to extreme support (modifications). Figure 4.2 shows the architectural structure of the ACT test's accessibility supports (note that the first level of support, embedded-universal supports, is identified as Levels 1–2 in Figure 4.2, depending on whether the support must be ordered in advance).

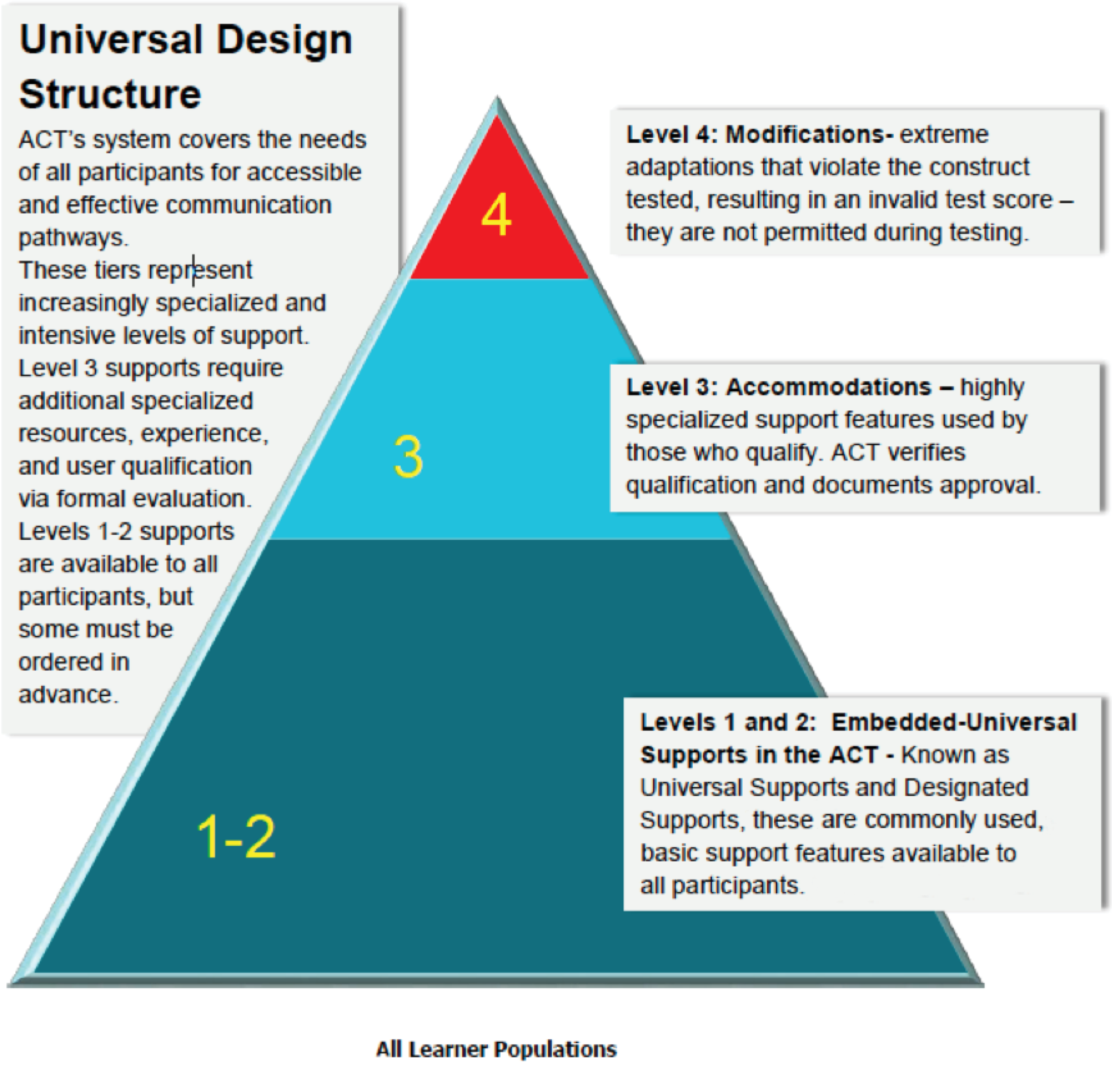


Figure 4.2 The ACT accessibility system structure.

The third level of support, modifications (identified as Level 4 in Figure 4.2), is not permitted in taking the ACT. The two permitted levels of support in the ACT accessibility system represent a continuum of supports, from least intensive to most intensive, and assume all users have communication needs that fall somewhere on this continuum. When an examinee has not requested any allowed accommodation-level supports, the system treats the examinee as a default user whose accessibility needs are sufficiently met through the embedded-universal test administration features represented by the base of the pyramid—that is, only the basic support features already embedded for all test takers (See Levels 1–2, "Embedded-Universal Supports" in Figure 4.2 and as described in the next section). The continuum of supports permitted in taking the ACT results in a personalized performance opportunity for all.

Support Levels 1–2: Embedded-Universal System Tools

Embedded-universal supports include system tools that meet the common, routine accessibility needs of the most typical test takers. All examinees are provided these tools as appropriate, even examinees that have no documented support plan. Embedded-universal system tools can be delivered in a fully standardized manner that is valid and include but are not limited to the following examples in online and paper tests:

- magnifier tool (online and paper)
- browser zoom magnification (online)
- test directions available on demand (online and paper)
- answer masking tool (online)
- line reader (online and paper)
- answer eliminator tool (online)
- keyboard navigation (online)
- scratch paper (online and paper)
- mark item for review (online and paper)
- We now offer color contrast and highligter as online accessibility tools

Embedded-universal system tools are common supports made available to all users upon launch or start of the test; they are the accessibility tools that nearly everyone uses routinely and assumes will be made available although test takers seldom think of them in this way. These tools are either embedded in the basic online test delivery platform or locally provided as needed. No formal request is needed for these supports, but some of these basic supports must be determined and planned for in advance of the test to ensure their availability.

Support Level 3: Allowed Accommodations

Allowed accommodations are available to users who qualify for a higher level of support. The ACT requires allowed accommodation-level supports to be requested by educational personnel on behalf of an examinee through the Test Accessibility and Accommodations (TAA) online system. This process allows any needed resources to be reviewed, approved, assigned with appropriate instructions for test administration, and documented for the examinee.

Typically, examinees who receive this level of support have a formally documented need and have therefore been locally identified as qualifying for—and have a written accommodations plan for—resources or equipment that requires expertise, special training, and/or extensive monitoring to select, administer, and even to use the support effectively and securely. These resources or equipment can include but are not limited to the following:

- braille EBAE, contracted, includes tactile graphics (paper)
- braille UEB with Nemeth contracted, includes tactile graphics (paper)
- braille UEB without Nemeth contracted, includes tactile graphics (paper)
- cued speech (paper)

- word-to-word bilingual dictionary, ACT-approved (online and paper)
- English audio USB, designed for user with blindness (paper)
- English audio reader script, designed for user with blindness (paper)
- signed exact English (SEE): test items (paper)
- abacus (paper)
- dictated responses (online and paper)
- extra time (online and paper)
- Breaks: supervised (online and paper) with each day (online and paper)
- keyboard or augmentative or assistive communication (AAC) + local print (online and paper)

Allowed accommodations are available to users who have been qualified by the local governing school or employment authority to use them, (for example: by a school district, or if the person has left school, by a work training agency, by an employer, or by a branch of military or other government service). Official determination of qualification for accommodation-level support by a governing school district or work authority is usually documented in writing in the form of an accommodation plan, or such qualification may have been routinely recognized and permitted for this person by that governing authority. The ACT requires examinees that use accommodation-level supports have a formally documented need, as well as relevant knowledge and familiarity with these supports. Accommodations must be requested 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.

Support Level 4: Modifications

Modifications are supports that are sometimes used during instruction to support learning, but when used in a testing situation, they do too much for the examinee that she or he is expected to do as an independent agent. In this way, modifications alter what the test is attempting to measure and thereby prevent meaningful access to performance of the construct that is being tested (see Figure 4.2). Because modifications violate the construct being tested, they invalidate performance results and communicate low expectations of examinee achievement. Modifications are not permitted in the ACT test.

Allowed Accommodations and Embedded-Universal Tools

As part of ACT's commitment to providing a fair testing experience for all examinees, the ACT test provides an integrated system of accessibility supports that include allowed accommodations as well as other forms (less intensive levels) of accessibility support. There are times when supports provided for those who test using the online format are combined with other types of locally provided or paper-format supports. The reverse is also true, as examinees using the paper format sometime also take advantage of certain online options. Regardless of test format, all examinees who use allowed accommodation-level accessibility features must have this use documented by appropriate school personnel. The general list of allowed assessment accessibility supports for the ACT is included in Tables 4.1–4.4. Full procedural requirements and instructions for using permitted supports during test administration are provided in the ACT test administration manuals.

4.3 Validity of Test Scores and Equal Opportunity to Benefit for All Examinees

ACT aims to ensure that all examinees may benefit equally from the ACT test. Allowed accommodations and embedded-universal accessibility supports administered under standardized conditions will result in a valid and fully reportable ACT score. Use of any accessibility supports that are not allowed or approved by ACT or not properly administered will violate what the test is designed to measure and will therefore result in a score that is invalid and noncomparable for the stated purposes of the test. Any scores that are produced in a way that would result in an invalid and noncomparable score for the stated purposes of national college reporting are treated as "non-college reportable" scores. This is true for any and all examinees who produce a score that in some way violates the constructs the ACT is designed to measure; therefore, that score will be noncomparable for the test's intended uses.

Tables 4.1–4.4 provide the list of allowed embedded-universal tools and allowed accommodation-level supports. As with any such list, there are circumstances where an individual need may be identified that has not been anticipated in the list of allowed supports. When this circumstance arises, ACT provides a mechanism, through the Test Accessibility and Accommodations (TAA) system, for the examinee to request consideration of this "other accommodation" (see last row of Table 4.4). When such a request occurs, documentation of qualification for use of accommodation-level supports will proceed as usual, and ACT will consult test design and content specialists to determine if the requested accommodation can be allowed. Through the TAA system, the examinee will be notified of the final determination.

4.4 The ACT 2017–18: Allowed Accessibility Supports for State & District Testing

Table 4.1 Presentation Supports

	Support Level		Content Area					Applies to:
Description	Paper	Online	Reading	English	Writing	Math	Science	State & District
Audio-Recording, Full Test	A	—	ˇ	ˇ	ˇ	ˇ	ˇ	SD
Reader Script, Full Test	A	—	ˇ	ˇ	ˇ	ˇ	ˇ	SD
Screen Reader	A	—	ˇ	ˇ	ˇ	ˇ	ˇ	SD
Text to Speech (available Spring 2018)	—	A	ˇ	ˇ	ˇ	ˇ	ˇ	SD
Translated Written Directions—12 Languages Provided (ELs) [6]	A	A	ˇ	ˇ	ˇ	ˇ	ˇ	SD
Translated Audio, Full Test (see note on p. 4.15) [1]	A [1]	A [1]	No	No	[1]	[1]	[1]	SD
Word-to-Word Dictionary (ELs) [6]	A	A	ˇ	ˇ	ˇ	ˇ	ˇ	SD
American Sign Language (ASL), Directions Only [7]	E [7]	A	ˇ	ˇ	ˇ	ˇ	ˇ	SD
Signed Exact English (SEE), Directions Only	A	A	ˇ	ˇ	ˇ	ˇ	ˇ	SD
Signed Exact English (SEE), Full Test	A	-	ˇ	ˇ	ˇ	ˇ	ˇ	SD
Cued Speech	A	-	ˇ	ˇ	ˇ	ˇ	ˇ	SD
English Braille American Edition (EBAE/ Nemeth), available with Tactile Graphics and Nemeth code for Math and Science (Contracted) Online support refers to required paper form companion to online test—see note. [2]	A [2]	A [2]	ˇ	ˇ	ˇ	ˇ	ˇ	SD
Unified English Braille (UEB), available with Tactile Graphics and Nemeth code for Math and Science (Contracted) [2]	A	A [2]	ˇ	ˇ	ˇ	ˇ	ˇ	SD
Tactile Graphics (stand-alone) with EBAE/ Nemeth [2]	A [2]	A [2]	—	—	—	ˇ	ˇ	SD

Table 4.1 Presentation Supports—continued

	Support Level		Content Area					Applies to:
Description	Paper	Online	Reading	English	Writing	Math	Science	State & District
Tactile Graphics (stand-alone) with UEB/ Nemeth [2]	A [2]	A [2]	—	—	—	˘	˘	SD
Large Print	A	—	˘	˘	˘	˘	˘	SD
Browser Zoom Magnification	—	E	˘	˘	˘	˘	˘	SD
Magnification	A	E	˘	˘	˘	˘	˘	SD
Line Reader (Straight-edge Tool, locally provided) [7] (Online tool, or Locally provided paper straight edge)	E [7]	E	˘	˘	˘	˘	˘	SD
Color Contrast (Online [3]) or Color Overlay (Locally Provided) [7]	E [7]	[3]	[3]	[3]	[3]	[3]	[3]	SD

Table 4.2 Interaction and Navigation Supports

	Support Level		Content Area					Applies to:
Description	Paper	Online	Reading	English	Writing	Math	Science	State & District
Abacus	A	A	—	—	—	˘	—	SD
Answer Masking Tool	E	E	˘	˘	˘	˘	˘	SD
Answer Eliminator Tool	E	E	˘	˘	˘	˘	˘	SD
Highlighter Tool [3]	A	3	3	3	3	3	3	SD
Keyboard Navigation	—	E	˘	˘	˘	˘	˘	SD
Use Test Booklet for Scratch Paper	E	—	˘	˘	˘	˘	˘	SD
Sheet of Paper to Use as Scratch Paper	A	E	˘	˘	˘	˘	˘	SD
Calculator, Including Accessible Calculator, all personally provided (headphones required for talking calculator) [4]	E	E	—	—	—	˘	—	SD

Table 4.3 Response Supports

	Support Level		Content Area					Applies to:
Description	Paper	Online	Reading	English	Writing	Math	Science	State & District
Respond in Test Booklet or on Separate Paper	A	—	˘	˘	˘	˘	˘	SD
Large Block Answer Sheet	A	—	˘	˘	˘	˘	˘	SD
Dictate Responses	A	A	˘	˘	˘	˘	˘	SD
Computer for Writing Essays and Constructed Responses	A	E	˘	˘	˘	˘	˘	SD
Speech to Text	A	A	˘	˘	˘	˘	˘	SD
Mark Item for Review Tool	E	E	˘	˘	˘	˘	˘	SD
Word Prediction External Device [5]	—	—	na	na	No[5]	na	na	SD

Table 4.4 General Test Conditions Supports

	Support Level		Content Area					Applies to:
Description	Paper	Online	Reading	English	Writing	Math	Science	State & District
Extra Time (ELs) [6]	A	A	˘	˘	˘	˘	˘	SD
Breaks	A	A	˘	˘	˘	˘	˘	SD
Multiple Days	A	A	˘	˘	˘	˘	˘	SD
Food or Medication for Individuals with Medical Need [7]	E [7]	E [7]	˘	˘	˘	˘	˘	SD
Special Seating/Grouping [7]	E [7]	E [7]	˘	˘	˘	˘	˘	SD
Location for Movement [7]	E [7]	E [7]	˘	˘	˘	˘	˘	SD
Individual Administration [7]	E [7]	E [7]	˘	˘	˘	˘	˘	SD
Administration at Optimum Time of Day [7]	E [7]	E [7]	˘	˘	˘	˘	˘	SD
Administration from Home or Care Facility [7]	E [7]	—	˘	˘	˘	˘	˘	SD
Separate Setting or Location (Familiar Setting and/or Small Group) (ELs) [6, 7]	E [7]	E [7]	˘	˘	˘	˘	˘	SD
Audio Amplification	A	A	˘	˘	˘	˘	˘	SD
Special Lighting [7]	E [7]	E [7]	˘	˘	˘	˘	˘	SD
Adaptive Equipment or Furniture [7]	E [7]	E [7]	˘	˘	˘	˘	˘	SD
Wheelchair Accessible Room	E [7]	E [7]	˘	˘	˘	˘	˘	SD
Personalized Auditory/Visual Notification of Remaining Time	A	A	˘	˘	˘	˘	˘	SD
Other Accommodations: Request Using TAA System	Yes	Yes	˘	˘	˘	˘	˘	SD

Table Key:

- Accommodation ("A" type) Supports used WITH required ACT approval listed in this table—WILL result in Reportable Score.
- Accommodation ("A" type) Supports used without required ACT approval, or not listed here (not allowed/not approved), will be assumed to be a Modification and will result in a NON-Reportable Score.
- Embedded Universal ("E" type) Supports listed in this table, if used in an otherwise proper administration—WILL result in a Reportable Score. Any examinee may use "E" type supports.
- The symbol "—" indicates this support is not applicable or not available to this test delivery format.

Footnotes from Tables on Pages 4.10–4.14:

[1] Provided ONLY as part of State & District negotiated contract for nonreportable score only.

[2] All users with blindness will need to use companion paper form braille/tactile graphics on math and science tests as critical interpretive information within math and science graphics will not be read aloud. This is required for both paper and online test formats.

[3] The online version of this support will be provided on all online tests when technology becomes available.

[4] Calculator use is not permitted for science test. Science test items requiring calculations are designed so that answering the items involves only minimal, rudimentary calculations. Some math-oriented science constructs that are assessed (e.g., recognizing relationships in scientific data, translation of data) are intended to be performed without use of graphing functionalities often present on calculators.

[5] The ACT writing test domain of Language Use & Conventions (including grammar, syntax and word usage) can be compromised by device usage. Reading, English, math and science are currently in MC format, making word prediction nonapplicable (na) at this time.

[6] **English Learners (ELs):** Four Accommodation-level ("A") supports now available to qualified ELs (verified by ACT per ESSA criteria) are indicated in the table.

[7] Embedded Supports (E), sometimes called "Local Arrangements," require prior planning and resource coordination at the local level to ensure proper, secure test administration.

4.5 2017 ACT Enhancements for English Learners

In 2016, ACT sought the counsel and advice of numerous K–12 and postsecondary education representatives, national researchers, and policy professionals who have expertise in identifying and serving English learners (ELs). Namely, ACT convened a blue-ribbon panel, conducted market research, and developed a robust internal research agenda to determine the impact on the ACT of providing supports to ELs. The panel carefully examined the potential impacts of each proposed support on construct validity, evaluated compliance with applicable federal and state laws regulating the ACT, and considered the impacts to test stakeholders in determining a fair test experience and delivery for all examinees, both those seeking supports and those testing under standard conditions. ACT acknowledges and appreciates the panel members' valuable contributions.

Beginning in the fall of the 2017–18 school year, ACT now provides supports on the ACT test to US EL students. These supports are limited to students who are enrolled in or qualified for a school district's EL program. As with all ACT accessibility goals, the goal of these supports is to ensure that the ACT scores earned by ELs accurately reflect what they have learned in school.

ACT adopted the following guiding principles for responding to requests from examinees identified as ELs for test supports:

1. Requirements and procedures for test supports must ensure fairness for all examinees, both those seeking supports and those testing under standard conditions.
2. Supports must be appropriate and reasonable for those with English Learner needs.
3. Documentation of English Learner status must meet established guidelines. Examinees must provide information about prior supports made in a similar setting, such as in academic classes and other testing situations.

ACT follows criteria delineated in federal law for establishing EL status, namely criteria identified in the Every Student Succeeds Act (ESSA). Therefore, to be eligible for supports on the basis of English learning needs, an examinee must establish, via submission of supporting documentation, that he or she is an individual:

- whose difficulties in speaking, reading, writing, or understanding the English language may be sufficient to deny the individual—
    - (i) the ability to meet the challenging State academic standards;
    - (ii) the ability to successfully achieve in classrooms where the language of instruction is English; or
    - (iii) the opportunity to participate fully in society;
- who is enrolled in an English Language program at a school located within the United States;
- who receives the requested supports on classroom tests via a formalized plan; and/or
- who provides results from an appropriate English language assessment that demonstrate the examinee's limited language proficiency.

Supporting documentation may include but is not limited to: an English Learner Plan, an Individualized Education Plan, other official support or accommodations plan, English language proficiency assessment results, and/or confirmation of eligibility or participation in an English language program.

All documentation submitted to ACT is kept confidential and is used solely to determine the applicant's eligibility for test supports. Test supervisors are also instructed to treat as confidential all information they receive relative to the examinee's EL status and testing supports.

EL supports are requested by schools on behalf of their students utilizing ACT's Test Accessibility and Accommodations (TAA) system. ELs may utilize one or all of the following supports if approved by ACT:

- extended time (not to exceed time and half)
- ACT-approved word-to-word bilingual dictionary (no definitions)
- written test instructions translated into written supplements in the student's native language.

  Note: After consulting with state officials and other sources, ACT will initially include the following limited number of languages: Spanish, Mandarin, Cantonese, Arabic, Russian, French, German, Vietnamese, Korean, Haitian Creole, Tagalog, and Somali. Test instructions will also be provided in these languages online
- verbal practice test instructions (provided online in English) may be translated by a local translator if the needed language is not already provided by ACT
- testing in a familiar environment/small group setting

Students who are certified as an EL and receive supports will receive a college-reportable score. ACT Score Reports do not include any specific information about the supports provided.

4.6 Accommodations, Score Validity, and Usage

Chapter 11 provides information on the evidence for validity of scores for those students with disabilities who test with accommodations and for English language learners. That chapter also provides available data on the incidence of the use of accommodations by various populations.

References

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

CAST. (2011) Universal Design for Learning Guidelines version 2.0. Wakefield, MA: Author. Retrieved from http://www.udlcenter.org/aboutudl/udlguidelines.

Fedorchak, G. (2012). Access by Design—Implications for equity and excellence in education. Whitepaper prepared for the Smarter Balanced Assessment Consortium.

Hansen, E. G., & Mislevy, R. J. (2007). Design Pattern for Improving Accessibility for Test Takers with Disabilities. ETS Research Report Series, 2008

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

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

Chapter 5

Test Administration

5.1 Overview

The ACT test must be uniformly administered 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, both in paper and online formats.

5.2 Administration Windows

The ACT is administered nationally and internationally on predetermined test dates. These dates and registration deadlines are available at www.act.org. The ACT is administered only on the days and at the times scheduled for a given test center. Tests administered on any other date or time, without prior approval from ACT Test Administration, will not be scored.

In addition to national and international test dates, the ACT is also administered during predetermined dates/windows for ACT State and District testing in the fall and spring.

5.3 Modes

Nationally, the ACT is administered on paper only. 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 12. In addition to standard formats, ACT offers accessibility supports and English learner (EL) supports for examinees approved for these accommodations. Additional information about accessibility and EL supports may be found in Chapter 4.

5.4 Locations

The ACT is administered at selected sites both nationally and internationally. Site locations are available at www.act.org. Typically, the sites consist of K–12 public, parochial, and private schools, and postsecondary institutions. To become a test center for national/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.

5.5 Policies and Procedures

5.5.1 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, protecting test security, and administering tests in a standardized manner. The manuals cover such things 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
- procedures for returning test materials to ACT

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

5.5.2 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 has committed, the test coordinator must notify ACT Test Administration 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.

5.5.3 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 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 answer questions in a group setting so everyone has the same information at the same time.

Chapter 6

Test and Information Security

6.1 Test Security

To ensure the validity of ACT test scores, test takers, 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, including the Test Security Principles and test security requirements. Those requirements are set forth in ACT's administration manuals and may be supplemented by ACT from time to time with additional communications to test takers and testing staff.

ACT's test security requirements are designed to ensure that examinees have an equal opportunity 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 are valid. Strict observation of the test security requirements is required to safeguard the validity of the results.

Testing staff must protect the confidentiality of the ACT test items and responses. Testing staff should be aware of and competent for 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 the scores. 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; or test proctor or test administrator misconduct (such as providing answers or questions to test takers or permitting test takers 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 the ACT test and the validity of its scores. These practices include (1) use of a reporting hotline through which individuals with information about misconduct on an ACT test can anonymously report such information to ACT, (2) data forensics to detect and respond to possible misconduct, and (3) web monitoring to detect testing misconduct, possible unauthorized disclosure of secure ACT test content, and other activity that might compromise the security of the ACT test or the validity of its scores.

6.2 Information Security

ACT's Information Security program framework is based on the widely recognized ISO/IEC 27000 standard (International Organization for Standardization, 2017). 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. The categories covered by the framework and brief statements of their importance to ACT are:

1. Information Security Program Management: This is overseen by the Information Security Officer at ACT. The Information Security Officer has responsibility 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 ISMS (Information Security Management System) guidelines defined in the 27005 standard. ACT also makes use of the SP NIST 800-37 Risk Assessment, which complies with FISMA (Federal Information Security Management Act) 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.

   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/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.
     - 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 to form a system of control to protect student data.

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 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 1) limit access to individuals who have a business need to know the information and 2) verify the individuals' identities. Access to ACT systems and data require authorization from the appropriate system owner. Active Directory, file permissions, and VPN (Virtual Private Network) remote access is administered by an Identity and Access management team who are 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 ASV (Approved Scanning Vendor). ACT has always maintained a "compliant" status in accordance with PCI-DSS (Data Security Standards) 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 intrusion detection strategies, tools, processes, and devices to look for unusual attack mechanisms and detect compromise of these systems. Network-based IDS sensors are deployed on Internet and extranet DMZ systems and networks, which provide alerting and procedures for review and response. Procedures include security review and approval of changes to configurations and 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 continuous scanning for rogue access points is performed.

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 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 response procedures, notification protocols, and escalation procedures. Forensics are performed at the direction of the Information Security Officer. ACT maintains a subscription service with a third party specializing in computer forensics in the event of a declared incident.

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: technical attacks (e.g., denial of service attack, malicious code attack, exploitation of a vulnerability), unauthorized behavior (e.g., unauthorized access to ACT systems, inappropriate usage of data, loss of physical assets containing confidential or confidential restricted data), or 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 Security Incident Response Team (ISIRT) with providing an around-the-clock (i.e., 24–7) coordinated security incident response throughout ACT. Information Security management has the responsibility and authority to manage the ISIRT and implement necessary ISIRP actions and decisions during an incident.

References

International Organization for Standardization. (2017). ISO/IEC 27000 family-information security management systems. Retrieved from https://www.iso.org/isoiec-27001-information-security.html.

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

Chapter 7

Scores, Indicators, and Norms

7.1 Overview

The ACT test contains four multiple-choice tests (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 due to the different purposes that they serve. They all contain not only scores of the student's overall performance on each subject test and across different subject tests but also detailed information about students' performance on specific areas within each subject test. Additional information is provided on the score reports to facilitate the interpretation of scores and for college and career planning.

The scores and indicators as well as the scoring process are introduced in Chapter 2. This chapter provides more information about the scores and indicators. Further information on the technical characteristics of the scores and indicators can be found in Chapters 9 and 10. Information reported on the score reports to facilitate college and career planning is described in Chapters 14 and 15.

7.2 Subject Test, Composite, STEM, and ELA Scores

The ACT student, high school, and college reports describe students' overall performance on the subject tests. Test scores on each subject as well as the Composite score and two combined scores are reported. The combined scores include the Science, Technology, Engineering, and Mathematics (STEM) score, a combination of the student's mathematics and science scores, and the English Language Arts (ELA) score, a combination of the student's English, reading, and writing scores for students who take the writing test. These scores constitute a major section of the score report as shown in Figure 7.1. Standard errors of measurement (SEMs), the ACT College Readiness Benchmarks, and national (US) and state ranks are also reported to facilitate the interpretation of these scores.

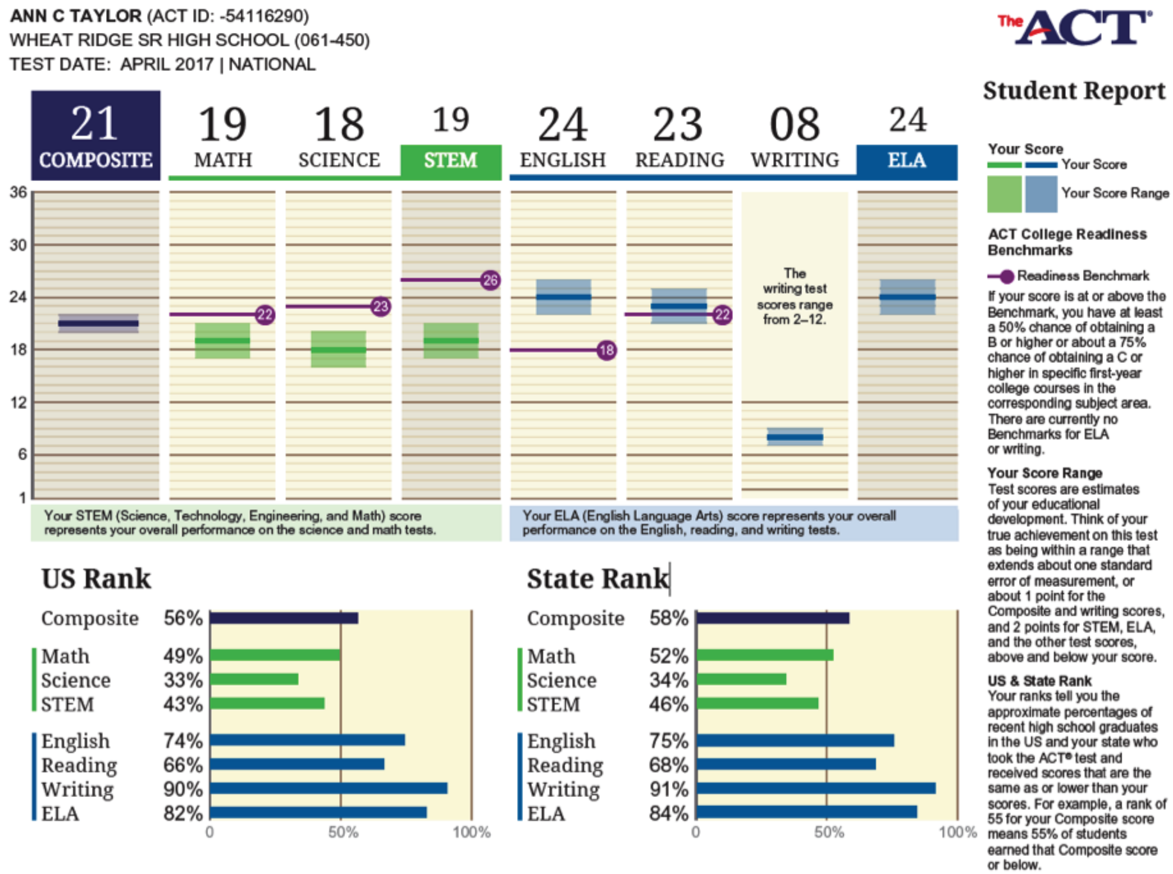


Figure 7.1 Overall score section of a sample ACT student score report

7.2.1 Subject Test Scores

Subject test scores are reported for the multiple-choice tests and the ACT writing test. For each of the four multiple-choice tests, the raw score is the number of correct responses. Raw scores are converted to scale scores through equating procedures to ensure that scores reported across test forms have a consistent meaning. Scale scores range from 1 to 36 for each of the four multiple-choice tests. Procedures for obtaining the 1–36 scale scores for the multiple-choice tests are described in Chapter 9.

For the ACT writing test, student responses are rated by two trained raters on four writing domains: Ideas & Analysis, Development & Support, Organization, and Language Use & Conventions. Detailed description of these domains can be found in Chapter 3. Using an analytic rubric, each rater assigns a score from 1 to 6 to each domain. Domain scores, ranging from 2 to 12, are the sum of the two rater scores. The writing test score is the average of the four domain scores rounded to the nearest integer. This writing score ranges from 2 to 12.

7.2.2 Composite, STEM, and ELA Scores

The ACT Composite score represents a student's overall performance on all the multiple-choice tests. It is the average of the four scale scores for English, mathematics, reading, and science rounded to the nearest whole number (fractions of 0.5 or greater round 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 with fractions rounded up. The ELA score represents a student's overall performance on the English, reading, and writing tests. For the calculation of ELA scores, the sum of the writing domain scores are converted to a scale of 1 to 36. Procedures for obtaining the 1–36 writing scale scores are described in Chapter 9. The ELA score is the rounded average of the English, reading, and the 1–36 writing scale scores. Only students who take the English, reading, and writing tests can receive an ELA score. The Composite, STEM, and ELA scores all range from 1 to 36. These scores are all comparable across different test forms.

7.2.3 Interpretation of the ACT Test Scores

The ACT score reports present additional information to help students and educators with the interpretation of the scores, including SEM, the ACT College Readiness Benchmarks, and the national and state ranks of the scores.

SEM and Score Ranges

The score report also contains information about the measurement precision of the subject test scores, and the Composite, STEM, and ELA scores. The SEMs are about one point for the writing and the Composite scores, and two points for the subject test and the STEM and ELA scores. Students' scores are reported with score ranges which are graphically represented by shaded areas around their scores. Detailed information about measurement precision is given in Chapter 10.

ACT College Readiness Benchmarks

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

Additional resources are available to facilitate the interpretation of ACT scores. ACT's College and Career Readiness Standards are sets of statements intended to help students, parents, and educators understand the meaning of test scores. The 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 Chapter 8 for more details about ACT's College and Career Readiness Standards and the ACT College Readiness Benchmarks.

Score Norms

The national (US) and state ranks can help students understand how their scores compare to other students in the nation or in their states. A rank shows the percentage of tested students whose scores are the same as or lower than a given student's score. ACT US and state ranks are based upon the most recent scores of high school seniors who graduated during the previous three years and tested in tenth, eleventh, or twelfth grade. The most recent US 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 tenth, eleventh, or twelfth grade, they are not intended to represent the performance of twelfth-grade students nationwide. See Appendix A for a description of the ranks for twelfth-grade students obtained from a 1995 nationally representative sample.

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

7.2.4 Summary Statistics, Effective Weights, and Correlations

Operational test data from five of the test forms administered in the 2015–2016 academic year were used to obtain descriptive statistics in this chapter. These data were based on large national samples. This section presents the summary statistics and correlational relationships among the subject test scores and the Composite, STEM, and ELA scores. Effective weights of each component in the Composite, STEM, and ELA scores are also reported.

Summary Statistics

The summary statistics of the ACT test scores are presented in Table 7.1.

Table 7.1 Summary Statistics of the ACT Test Scores

Statistic	English	Mathematics	Reading	Science	Composite	STEM	Writing	ELA
Mean	21.63	21.65	22.50	21.85	22.03	22.00	7.22	22.13
SD	6.38	5.36	6.11	5.14	5.22	4.97	1.66	5.39
Skewness	0.15	0.40	0.18	0.22	0.22	0.33	0.00	-0.09
Kurtosis	-0.57	-0.67	-0.69	0.10	-0.61	-0.41	0.19	-0.58

Effective Weights

The Composite, STEM, and ELA scores are the rounded averages of the subject test scores. Specifically, the English, mathematics, reading, and science test scale scores are weighted equally to form the Composite score, the mathematics and science scale scores are weighted equally to form the STEM score, and the English, reading, and writing scale scores are weighted equally to form the ELA score. Forming scores in such a way indicates that for the ACT Composite, STEM, and ELA scores, the weights used in the calculation are ¼, ½, and ⅓, respectively, and they are often referred to as *nominal weights*.

Other definitions of the contribution of a test score to a combined score are also available. *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). To obtain effective weights, score covariances are first obtained. The effective weight for a test can be calculated by summing the values in the appropriate row and dividing the resulting value by the sum of all covariances among the tests using the formula

$$(\textit{effective weight})_x = \frac{\sum_y \text{cov}_{xy}}{\sum_x \sum_y \text{cov}_{xy}} ,$$

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

Taking the Composite score as an example, to obtain effective weights for the four multiple-choice tests, scale score covariances from one test form administered in the 2015–2016 academic year were computed (see Table 7.2). The effective weight for the English test was computed by adding up the four numbers in the first row (42.10, 27.52, 31.79, and 25.57). 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 were obtained in a similar fashion.

Table 7.3 shows the ranges of effective weights for the Composite, STEM, and ELA scores based on five of the test forms administered in the 2015–2016 academic year. For these scores, the effective weights were fairly stable across the five 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. For the STEM score, the mathematics scores had larger weights than the science scores because the mathematics scores had larger score variances than science. Note that these effective weights were from the large national samples and that the weights might differ considerably from those for other examinee groups.

Table 7.2 Scale Score Covariances for Multiple-Choice Tests from One ACT Test Form

Test	English	Mathematics	Reading	Science
English	42.10	27.52	31.79	25.57
Mathematics	27.52	29.37	23.13	21.56
Reading	31.79	23.13	36.89	23.14
Science	25.57	21.56	23.14	24.52

Table 7.3 Range of Effective Weights of the ACT Tests

	Range of Effective Weights		
Test	Composite	STEM	ELA
English	0.28–0.29		0.36–0.39
Mathematics	0.23–0.23	0.50–0.53	
Reading	0.26–0.27		0.35–0.37
Science	0.22–0.23	0.47–0.49	
Writing			0.24–0.29

Correlations

Table 7.4 shows the median correlations among the ACT test scores based on operational data from five of the test forms administered in the 2015–2016 academic year. 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 other scores. Score reliability of the ACT tests can be found in Chapter 10.

Table 7.4 Correlations among the ACT Test Scores

Score	English	Mathematics	Reading	Science	Composite	STEM	Writing	ELA
English	1.00	0.76	0.81	0.77	0.93	0.81	0.58	0.93
Mathematics		1.00	0.69	0.78	0.88	0.94	0.50	0.75
Reading			1.00	0.75	0.90	0.75	0.54	0.91
Science				1.00	0.90	0.94	0.51	0.78
Composite					1.00	0.95	0.59	0.93
STEM						1.00	0.54	0.81
Writing							1.00	0.78
ELA								1.00

7.3 Detailed Results

As shown in Figure 7.2, detailed results are provided for students' performance on finer skills and domains within each subject test: reporting category scores and ACT Readiness ranges for each multiple-choice test as well as domain scores for the ACT writing test.

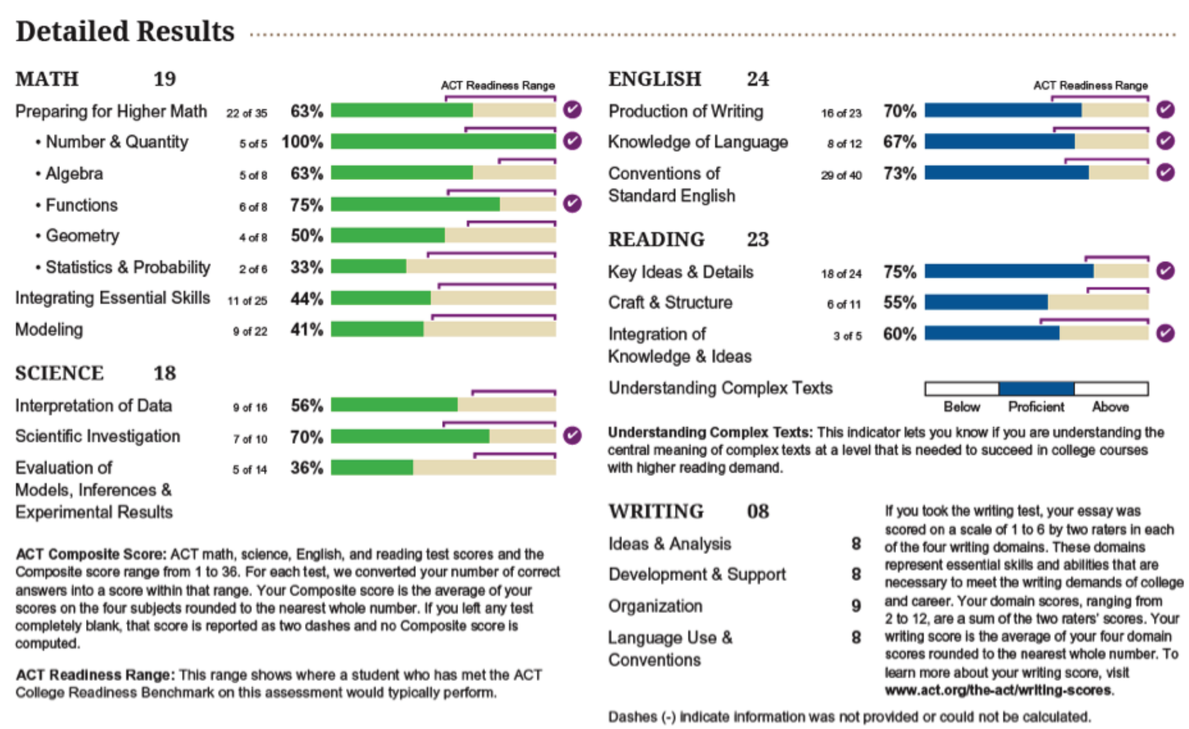


Figure 7.2 Detailed results section of a sample ACT student score report

7.3.1 Reporting Categories and ACT Readiness Ranges

ACT reporting categories are aligned with the ACT College and Career Readiness Standards and other standards that target college and career readiness. Scores on items that measure similar skills are grouped together to provide students with more detailed information within each subject. There are three reporting categories each for English, reading, and science, and eight for mathematics. These reporting categories make it easier for students, parents, and educators to gain insight into students' performance by better understanding students' strengths and areas for improvement on each subject. The reporting category scores replaced the subscores (e.g., Intermediate Algebra/Coordinate Geometry) that were previously reported.

The number of items for a particular reporting category can vary across different test forms. 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 reported. In addition, for each reporting category, there is an ACT Readiness range indicating the expected percent correct scores for students who are at or above the ACT College Readiness Benchmark for that specific subject.

ACT student data are used to create a predictive relationship between the ACT College Benchmark on the overall subject test and each of the test's reporting categories. For example, a Readiness range is developed for each of the three English reporting categories. For the first reporting category, Production of Writing, student scores on the overall English test and scores on the Production of Writing reporting category are used to estimate the predictive relationship between the two scores through linear regression. This relationship is then used to identify the minimum percent correct value for the reporting category that corresponds to the Benchmark on the overall English test. Students with percent correct values at or above the minimum percent correct value obtained during this process are considered to be within the ACT Readiness range. The maximum on the ACT Readiness ranges 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 as well as the reporting categories of the other multiple-choice tests.

Information about the development and blueprints of ACT reporting categories can be found in Chapter 3, and details about the interpretation of ACT reporting categories and ACT Readiness ranges can be found in the *ACT Reporting Category Interpretation Guide* by Powers, Li, Suh, and Harris (2016).

7.3.2 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 measure 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 on each of the four domains. If the scores from the two raters differ by more than one score point on any of the domains, a third rater evaluates the essay and resolves 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 for scoring the writing test can be found in Chapter 3.

Table 7.5 presents the summary statistics of writing domain scores and the overall writing scores based on five of the writing test forms administered in the 2015–2016 academic year. Table 7.6 presents the correlations among these scores.

Table 7.5 Summary Statistics of the ACT Writing and Writing Domain Scores

Statistic	Ideas & Analysis	Development & Support	Organization	Language Use & Conventions	Writing Score
Mean	7.17	6.88	7.07	7.42	7.22
SD	1.72	1.73	1.68	1.63	1.66
Skewness	-0.01	0.03	-0.06	0.00	0.00
Kurtosis	0.15	0.00	0.18	0.31	0.19

Table 7.6 Correlations among the ACT Writing and Writing Domain Scores

Score	Ideas & Analysis	Development & Support	Organization	Language Use & Conventions	Writing Score
Ideas & Analysis	1.00	0.95	0.95	0.93	0.98
Development & Support		1.00	0.95	0.90	0.95
Organization			1.00	0.93	0.97
Language Use & Conventions				1.00	0.96
Writing Score					1.00

7.3.3 Understanding Complex Texts Indicator

The Understanding Complex Texts (UCT) indicator is reported to show whether students understand the central meaning of complex texts at a level that is 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' performance on these items are classified into three levels: Below Proficient, Proficient, and Above Proficient.

The performance levels were established through a special study that linked students' scores on the UCT items to their college course grades (Allen, Bolender, Fang, Li, & Thompson, 2016). This special study examined the UCT scores and course grades of 263,265 students from 439 postsecondary institutions. To obtain UCT scores for the study, the UCT test items were classified 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 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 using UCT scores was used to predict 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. For each course and institution, the UCT score associated with a 50% chance of earning a B or higher grade was identified. These results were aggregated over a weighted sample of institutions to find the Proficient cut score of 9. 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 was obtained 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 two standard errors of measurement above the Proficient cut score. For additional information on the development of the UCT cut scores, see the full report by Allen et al. (2016). More information about technical characteristics of the UCT indicator can be found in Chapter 10.

7.3.4 Progress Toward the ACT National Career Readiness Certificate Indicator

The Progress Toward the ACT National Career Readiness Certificate (ACT NCRC) indicator is based on students' ACT Composite scores. This indicator provides an estimate of students' likely performance on the ACT WorkKeys National Career Readiness Certificate. The WorkKeys NCRC 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 the ACT WorkKeys Assessments. Scores on the ACT WorkKeys Assessments determine the certificate level—no certificate, Bronze, Silver, Gold, or Platinum—an individual can earn. 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 http://workforce.act.org/credential.

Data from nearly 79,000 11th and 12th graders who took the ACT and all three WorkKeys NCRC assessments in the 2017–2018 academic year were used to establish a link between minimum ACT Composite scores and the WorkKeys NCRC levels (Radunzel & Fang, 2018). These students had taken the refreshed ACT WorkKeys NCRC assessments released in 2017. Logistic regression was used to identify the ACT Composite cut 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 (e.g., 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 NCRC indicator classifies students into one of five levels[1]: Unlikely to earn an WorkKeys NCRC (below 13), likely to earn a Bronze Level NCRC (13–16), likely to earn a Silver Level NCRC (17–21), likely to earn a Gold Level NCRC (22–26), and likely to earn a Platinum Level NCRC (27–36).

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

References

ACT. (2014). *ACT Technical Manual*. Iowa City, IA: Author.

Allen, J. (2013). *Updating the ACT College Readiness Benchmarks* (ACT Research Report No. 2013-6). Iowa City, IA: ACT, Inc.

Allen, J., Bolender, B., Fang, Y., Li, D., & Thompson, T. (2016). *Relating the ACT Indicator Understanding Complex Texts to College Course Grades* (ACT Research & Policy Technical Brief). Iowa City, IA: ACT, Inc.

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

Powers, S., Li, D., Suh, H., & Harris, D. J. (2016). *ACT Reporting Category Interpretation Guide: Version 1.0* (ACT Working Paper No. 2016-5). Iowa City, IA: ACT, Inc.

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

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

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

Chapter 8

ACT's College and Career Readiness Standards and College Readiness Benchmarks

8.1 Overview

This chapter describes ACT's College and Career Readiness Standards and College Readiness Benchmarks. The focus of this chapter is to provide background on the standards and benchmarks–e.g., their purpose, how they are developed and maintained, and how to interpret them.

The 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
- 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.

8.2 ACT's College and Career Readiness Standards

8.2.1 Description of the College and Career Readiness Standards

In 1997, ACT began an effort to make the ACT test results more informative and useful. This effort yielded ACT's College and Career Readiness Standards. The College and Career Readiness Standards are statements that describe what students who score in various score ranges on the tests are likely to know and to 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," while 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." The Standards reflect a progression of skills in each of the five tests: English, mathematics, reading, science, and writing. ACT has organized the standards by *strands*—related areas of knowledge and skills within each test—for ease of use by teachers and curriculum specialists. The complete College and Career Readiness Standards are presented at the end of this chapter and posted on ACT's website: www.act.org. They also are available in poster format. To order additional posters, please email customerservices@act.org. ACT also offers College and Career Readiness Standards Information Services, a supplemental reporting service based on the Standards.

College and Career Readiness Standards for the ACT are provided for six score ranges (13–15, 16–19, 20–23, 24–27, 28–32, and 33–36) along a score scale of 1–36. 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 able to demonstrate most or all of the skills and understandings in the 13–15, 16–19, and 20–23 score ranges.

College and Career Readiness Standards for the writing test, which ACT developed in 2005 and updated with enhancements in 2015, are available only for the ACT test and are provided for five score ranges (3–4, 5–6, 7–8, 9–10, and 11–12) in four writing domains, based on ACT writing test scores attained (the sum of two raters' scores using the six-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.

8.2.2 Determining the Score Ranges for the 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. This information was considered within 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 has conducted over the last 40 years. ACT's Course Placement Service provides colleges and universities with cutoff scores that are 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 of all four tests.

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 in relating the ACT multiple-choice test scores to students' skills and understandings.

8.2.3 Developing the College and Career Readiness Standards

After reviewing the normative data, college admissions criteria, and information obtained through ACT's Course Placement Service, content area test specialists (highly qualified subject-matter experts in each area) wrote the College and Career Readiness Standards based on their analysis of the knowledge and skills 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 College and Career Readiness Standards a high degree of confidence that students scoring in 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 along with tables that showed the percentages of students in each score range who answered each test item correctly (i.e., item difficulty). Item difficulties were computed separately based on groups of students whose scores fell within each of the defined score ranges.

Each content team was provided with ten forms of the ACT test and the item difficulties computed separately for each score range for each of the items on the forms. For example, the mathematics content team reviewed ten 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 8.1.

The shaded areas in Table 8.1 show the items that met the .80-or-above item difficulty criterion for each of the score ranges. As illustrated in Table 8.1, a cumulative effect can be noted: the items that are correctly answered by 80% of the students in Score Range 16–19 also appear in Score Range 20–23; the items that are correctly answered by 80% of the students in Score Range 20–23 also appear in Score Range 24–27; 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 8.2 reports the total number of test items reviewed for each content area.

The procedures above allowed the content teams to conceptualize what is measured by each of the ACT tests. Specifically, each content team followed the same process as they reviewed the test items in each multiple-choice test of the ACT. Below are the detailed steps.

1. Multiple forms of each test were distributed.
2. The knowledge, skills, and understandings that are necessary to answer the test items in each score range were identified.

Table 8.1 Illustrative Listing of Mathematics Item Difficulties by Score Range

	Score range					
Item no.	13–15	16–19	20–23	24–27	28–32	33–36
1	.62	.89	.98	.99	1.00	1.00
2		.87	.98	.99	.99	1.00
6	.60	.86	.94	.97	.99	.99
7	.65	.92	.98	.99	.99	1.00
20		.84	.94	.97	.98	.99
27		.85	.97	.99	.99	.99
4			.92	.97	.99	1.00
5			.94	.97	.99	.99
.			.	.	.	.
.			.	.	.	.
.			.	.	.	.
8			.82	.95	.98	.99
9			.80	.89	.96	.99
21			.82	.92	.97	.99
13				.90	.97	.99
15				.90	.97	.99
17				.87	.98	1.00
18				.83	.93	.98
22				.81	.91	.98
24				.83	.96	.98
29				.87	.98	1.00
34				.86	.95	.99
36				.82	.93	.99
39				.85	.96	.99
44				.84	.96	.99
25					.95	.99
28					.97	1.00
.					.	.
.					.	.

Table 8.1 Illustrative Listing of Mathematics Item Difficulties by Score Range—continued

	Score range					
Item no.	13–15	16–19	20–23	24–27	28–32	33–36
.					.	.
35					.86	.96
47					.86	.97
32						.95
33						.92
46						.90
49						.95
51						.98
52						.98
53						.92
56						.98
57						.86
58						.95
59						.86
60						.96

Table 8.2 Number of ACT Items Reviewed During 1997 National Review

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

3. The *additional* knowledge, skills, and understandings that are necessary to answer the test items in the *next* score range were identified. This process was repeated for all the 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 larger group of content specialists.
5. The composite list was reviewed by each content specialist, and ways to generalize and to consolidate the various skills and understandings were identified.
6. The content specialists met as a group to discuss the individual, consolidated lists and prepared a master list of skills and understandings, 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 necessary.
8. The adjustments were reviewed by the content specialists and "final" revisions were made.
9. The "final" list of skills and understandings was used to review additional test forms. The purpose of this review was to determine whether the College and Career Readiness Standards adequately and accurately described the skills and understandings measured by the items specific to each score range.
10. The College and Career Readiness Standards were once again refined.

These steps were used to review test items for all four multiple-choice tests.

Conducting an independent review of the 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 College and Career Readiness Standards. These teachers and researchers were asked to provide ACT with independent, authoritative reviews of the College and Career Readiness Standards.

The content area experts were selected from among candidates having experience with and an understanding of the academic tests on the ACT. 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 College and Career Readiness 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 understandings across the score ranges. Each national content area team consisted of three college faculty members currently teaching courses in curriculum and instruction, and three classroom teachers, one each from Grades 8, 10, and 12. The reviewers were provided with the complete set of College and Career Readiness Standards and a sample of test items falling in 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 0.05. That is, the standard error of the estimated percent of matches to the Standards should be no greater than 0.05. 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 College and Career Readiness Standards. Each reviewer received a packet of materials that contained the College and Career Readiness Standards, sets of randomly selected items (17 per score range), introductory materials about the College and Career Readiness 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 understandings from the lowest score range to the highest score range. The reviewers were asked to record their ratings using a five-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 the Standards 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 8.3. 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 all the national consultants. The teams reviewed all suggestions and adopted a number of helpful clarifications in the language of the Standards, particularly in the language of the ACT reading test Standards—in which the original language had failed to meet the target criterion.

Table 8.3 Percentage of Agreement of 1997 National Expert Review

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

8.2.4 The College and Career Readiness Standards for Writing

In 2005, the College and Career Readiness Standards for Writing were developed. Following the enhancements to the ACT writing test in 2015, the Standards were updated. These Standards are statements of what students who score in various ranges on the ACT writing test are likely to be able to do. College and Career Readiness Standards for writing are provided across four domains for five writing test score ranges: 3–4, 5–6, 7–8, 9–10, and 11–12.

The score ranges and the College and Career Readiness Standards for the ACT writing test were derived from the ACT writing test scoring rubric. The writing test scoring rubric is a four-domain, six-point descriptive scale to which writing essays are compared in order to determine their scores. 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 College and Career Readiness Standards for Writing, which were derived from a holistic scoring rubric. The design of the enhanced writing test and accompanying College and Career Readiness 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 College and Career Readiness Standards for Writing, 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 3–4, 5–6, 7–8, 9–10, and 11–12 would best distinguish students' levels of writing achievement so as to assist teachers, administrators, and others in relating ACT test scores to students' skills and understandings. Writing that receives a score below 3 does not permit useful generalizations about the student's writing abilities in that domain.

8.2.5 Periodic Review of the College and Career Readiness Standards

ACT periodically conducts internal reviews of the College and Career Readiness Standards. ACT identifies three to four new forms of the ACT, and then analyzes the data and the corresponding test items specific to each score range. Topics are also compared to data from the most recent ACT National Curriculum Survey (e.g., ACT, 2016a). The purposes of these reviews are to ensure that (a) the Standards reflect the most important knowledge and skills for college and career readiness, (b) the Standards reflect what is being measured by the items in each score range, and (c) the Standards reflect a cumulative progression of increasingly sophisticated skills and understandings from the lowest score range to the highest score range. Minor refinements intended to update and clarify the language of the Standards have resulted from these reviews.

8.2.6 Interpreting and Using the College and Career Readiness Standards

Because new ACT test forms are developed at regular intervals and because no one test form measures all of the skills and knowledge included in any particular standard, the College and Career Readiness Standards must be interpreted as knowledge and skills that most students who score in 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 cautiously.

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 College and Career Readiness Standards should be interpreted in a responsible way that will help students, parents, teachers, and administrators to 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 post high school plans
- Enhance the communication between secondary and postsecondary institutions
- Identify the knowledge and skills students entering their first year of postsecondary education should know and be able to do in the academic areas of language arts, mathematics, and science
- Assist students as they identify skill areas they need to master in preparation for college-level coursework

8.3 ACT's College Readiness Benchmarks

8.3.1 Description of the College Readiness Benchmarks

The ACT College Readiness Benchmarks are scores on the ACT subject tests 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 (see Table 8.4). For example, the ACT English Benchmark corresponds to a minimum score of 18 on the ACT English test and is derived based on course success in English Composition I.

Table 8.4 ACT College Readiness Benchmarks

College course(s) or course area	ACT test score	The ACT Benchmark
English Composition I	English	18
College Algebra	Mathematics	22
Social science courses	Reading	22
Biology	Science	23
Calculus I, Biology, Chemistry, Physics, and Engineering	STEM	26
English Composition I and social science courses	ELA	20

Note. Social science courses included American History, Other History, Psychology, Sociology, Political Science, and Economics. The ACT STEM score is the rounded average of the ACT mathematics and science test scores. The ACT ELA score is the rounded average of the ACT English, reading, and writing test scores.

The ACT College Readiness Benchmarks are empirically derived based on the actual performance of students in college. As part of its research services, ACT provides reports to colleges to help them place students in entry-level courses as accurately as possible. In providing these research services, ACT has an extensive database consisting of course grade and test score data from a large number of first-year students and across a wide range of postsecondary institutions. These data provide an overall measure of what it takes to be successful in selected first-year college courses. The numbers and types of colleges vary by course. Because these colleges constitute a "convenience" sample (i.e., based on data from colleges that chose to participate in ACT's research services), there is no guarantee that it is representative of all colleges in the United States. Therefore, ACT applies weights when combining the results across institutions to obtain the Benchmarks to ensure that the sample of institutions represents the population of institutions attended by ACT-tested students in terms of college type (2-year and 4-year) and selectivity.

Three separate studies were conducted to develop the ACT College Readiness Benchmarks. The first developed the ACT Benchmarks in English, reading, mathematics, and science. The second developed the STEM Readiness Benchmark, and the third developed the ELA Readiness Benchmark. These three studies are described in the next sections.

8.3.2 Development of ACT's English, Mathematics, Reading, and Science College Readiness Benchmarks

In the spring of 2003, Allen and Sconing (2005) conducted a study to establish readiness benchmarks for common first-year college courses based on ACT scores. Benchmarks were developed for the following courses or course combinations: English Composition I, using the ACT English score; College Algebra, using the ACT mathematics score; Biology, using the ACT science score; and a combination of six social science courses, using the ACT reading score (see Table 8.4). The ACT College Readiness

Benchmarks were updated in 2013 using data from more recent high school graduates (Allen, 2013). As such, the Benchmarks are subject to change over time. Some of the possible reasons for reevaluating and updating the Benchmarks from time to time include a change in college grading standards, an aggregate change in college student performance, and a change in the level of alignment of secondary and postsecondary course content.

Data and method. Data for the most recent study (Allen, 2013) came from colleges or groups of colleges that participated in ACT's research services, including the Course Placement Service and Prediction Service. Results were based on 96,583 students from 136 colleges for English Composition I, 70,461 students from 125 colleges for College Algebra, and 41,651 students from 90 colleges for Biology. Six different courses were considered for the social science analyses: American History, Other History, Psychology, Sociology, Political Science, and Economics. Results for the social science courses were based on 130,954 students from 129 colleges.

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 .50. According to Sawyer (1989), this score point most accurately classifies the group into those who would be successful and those who would not. 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 were determined based on the median cutoff scores across colleges. For further details of the research methods, see Allen (2013).

Results. Table 8.5 gives the median ACT cutoff scores across colleges, along with the first and third quartiles. Scores of 18 for English, 22 for College Algebra, 22 for social science, and 23 for Biology represent ACT Benchmarks that would give a student at a typical college a reasonable chance of success in these courses; that is, at least a 50% chance of earning a B or higher grade. Moreover, these cutoff scores were associated with a 73% to 79% chance of earning a C or higher grade.

For the 2016 ACT-tested graduating class, 61% of students met the ACT Benchmark in English, 41% met the ACT Benchmark in mathematics, 44% met the ACT Benchmark in reading, 36% met the ACT Benchmark in science, and 26% met all four Benchmarks (Table 8.6; ACT, 2016d). The corresponding percentages for ACT-tested, first-year, and full-time college enrollees in 2015–2016 were higher by 13 to 16 percentage points (ACT, 2016c).

Summary. Students, parents, and counselors can use the Benchmarks to determine the academic areas in which students are ready for college course work and areas in which they need improvement. Although the Benchmarks are useful predictors of success in first-year college courses, ACT scores above the cutoffs do not guarantee success since factors other than academic preparedness, such as motivation and good study habits, are also important for success in college (Mattern et al., 2014).

Table 8.5 ACT College Readiness Benchmarks by Subject

Course	ACT test	Median score[a]	1st Quartile/3rd Quartile
English Composition I	English	18	16/20
College Algebra	Mathematics	22	21/24
Social science	Reading	22	20/24
Biology	Science	23	22/25

[a]The College Readiness Benchmarks were determined based on the median cutoff scores across colleges.

Table 8.6 Percentage of Students Meeting the ACT College Readiness Benchmarks, 2015–2016

ACT Benchmark	High school graduating class	Enrolled college freshmen[a]
English	61	77
Mathematics	41	54
Reading	44	57
Science	36	49

[a]Enrollment based on National Student Clearinghouse data.

8.3.3 Development of the ACT STEM Readiness Benchmark

In fall 2015, ACT introduced a STEM score for the ACT test that provides students and educators with more insight into critical aspects of students' readiness for first-year college course work in science, technology, engineering, and mathematics (STEM) disciplines. The STEM score is the rounded average of the ACT mathematics and science test scores and represents students' overall performance in these subjects. A study by Mattern, Radunzel, and Westrick (2015) suggested that academic readiness for STEM coursework may require higher scores than those indicated by the ACT College Readiness Benchmarks given that Calculus instead of College Algebra appears to be the typical first mathematics course of students majoring in STEM fields. Typical first science courses taken by students majoring in STEM fields included Chemistry, Biology, Physics, and Engineering. In a subsequent study, Radunzel, Mattern, Crouse, and Westrick (2015) identified the ACT STEM score associated with a reasonable chance of success in first-year mathematics and science courses taken frequently by STEM majors.

Data and method. Data used to develop the ACT STEM Readiness Benchmark based on the ACT STEM score came from four-year postsecondary institutions that participated in research services offered by ACT and included students from the 2005 through 2009 freshman cohorts. Results were based on nearly 85,000 students from 78 institutions. The same methodology as the individual subject area ACT College

Readiness Benchmarks was used to develop the ACT STEM Readiness Benchmark (Allen, 2013; Mattern et al., 2015). Briefly, the grades earned in first-year STEM courses (Calculus, Biology, Chemistry, Physics, and Engineering) were combined in a single course-success model to determine the ACT STEM test score that was associated with at least a 50% chance of earning a B or higher grade in those courses. Hierarchical logistic regression was used to model the probability of success in a course within each college as a function of the ACT STEM score. The model also included an indicator for the content area (mathematics versus science). Typical probabilities of success by the ACT STEM score were determined by calculating the median probabilities across institutions within each content area and then averaging the probabilities across the two content areas giving equal weight to the two areas.

Results. When combining grade data for Calculus and multiple science courses into a single course-success model, 26 was the ACT STEM score associated with at least a 50% chance of earning a B or higher grade in a STEM-related course (Figure 8.1). Moreover, this cutoff score was associated with an approximate 75% chance of earning a C or higher grade. The ACT STEM score of 26 also corresponded to the average of the ACT mathematics (27) and science (25) scores, which were derived by using separate STEM content area course-success models for Calculus and a combination of science courses (Mattern et al., 2015).

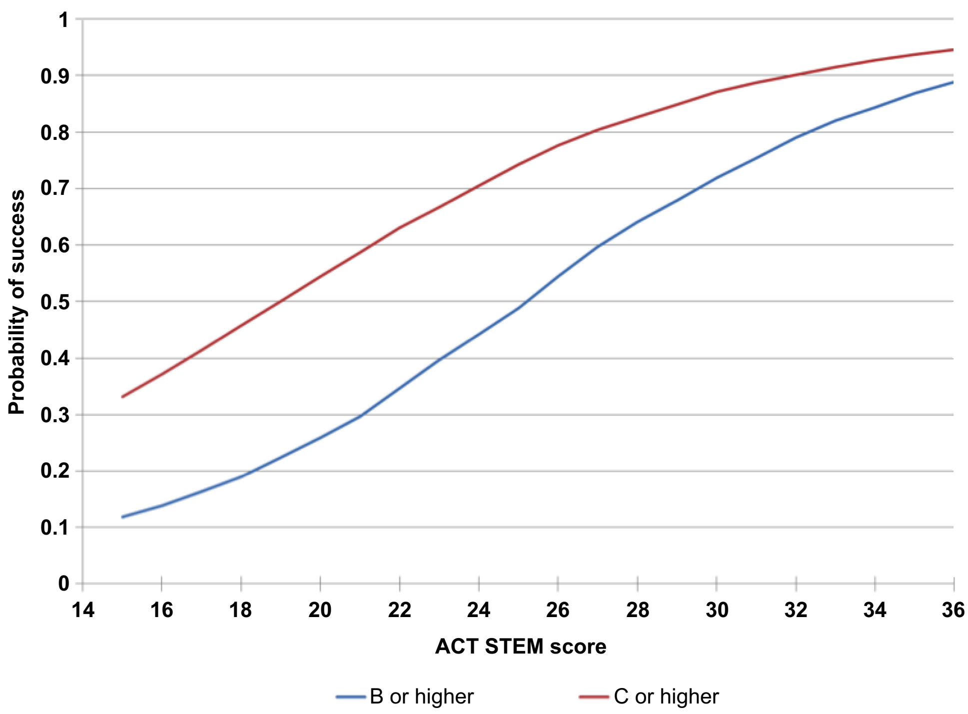


Figure 8.1 The typical probability of success in STEM-related courses by the ACT STEM score. The mathematics course was Calculus I. The science courses included Biology, Chemistry, Physics, and Engineering.

Summary. The ACT STEM Readiness Benchmark can be used to help gauge overall student readiness for STEM-related coursework. Based on the ACT STEM Readiness Benchmark of 26, only 20% of the 2016 ACT-tested high school graduating class was ready for first-year STEM-related college courses.

8.3.4 Development of the ACT ELA Readiness Benchmark

To provide students with an aggregate measure of their readiness in English, reading, and writing, ACT introduced the ACT ELA score in fall 2015 for students who take the optional ACT writing test. The ACT ELA score is the rounded average of the ACT English, reading, and writing scores; it ranges from 1 to 36. Given the importance of integrated literacy skills for academic and workplace success (Camara, O'Connor, Mattern, & Hanson, 2015), Radunzel, Westrick, Bassiri, and Li (2017) explored ELA readiness and what that means in relation to being successful in first-year ELA-related courses in English and the social sciences. The ELA-related courses commonly taken during the first-year were English Composition I, American History, Other History, Psychology, Sociology, Political Science, and Economics. This pattern of ELA-related course taking was observed irrespective of students' general major categories, including being observed among students from more specific ELA-related majors. These are the same courses used to derive the separate ACT College Readiness Benchmarks in English and reading (Allen, 2013). The study by Radunzel et al. identified the ACT ELA score associated with a reasonable chance of success in these seven first-year, ELA-related English and social science courses.

Data and method. Data used to develop the ACT ELA Readiness Benchmark came from 233 two- and four-year postsecondary institutions that participated in research services offered by ACT and included 198,275 students from the 2006 through 2014 freshman cohorts who had taken the former ACT writing test. Using a concordance table, students' ACT writing scores were converted to current ACT writing scores that are used to calculate the ACT ELA score (ACT, 2015). Students' ELA scores were estimated as the rounded average of the ACT English, reading, and concorded writing scores from the student's latest test record when the student took the ACT with writing; see Appendix A from the full research report by Radunzel et al. (2017) for empirical evidence supporting the use of the concorded writing scores in calculating an ACT ELA score for earlier cohorts to be used in the development of a preliminary ACT ELA Benchmark.

The same methodology as the individual subject area ACT College Readiness Benchmarks was used to develop the ACT ELA Readiness Benchmark (Allen, 2013; Mattern et al., 2015). Briefly, the grades earned in seven courses in English and the social sciences commonly taken during the first year (English Composition I, American History, Other History, Psychology, Sociology, Political Science, and Economics) were combined in a single course-success model to determine the ACT ELA test score associated with at least a 50% chance of earning a B or higher grade in those courses. For students who were enrolled in multiple ELA-related courses during the same term, grade information for a single course was randomly selected for inclusion in the analyses. Hierarchical logistic regression was used to model the probability of success in a course as a function of the ACT ELA score within each college. The model also included an indicator for content area (English versus the social sciences). Typical probabilities of success by the ACT ELA score were determined by calculating the median probabilities across institutions within each content area and then averaging the probabilities across the two content areas giving equal weight to the two areas.

Results. When combining grade data for English Composition I and multiple social science courses into a single course-success model, 20 was the ACT ELA score associated with at least a 50% chance of earning a B or higher grade in an ELA-related course (Figure 8.2). This cutoff score was also associated with an approximate 75% chance of earning a C or higher grade.

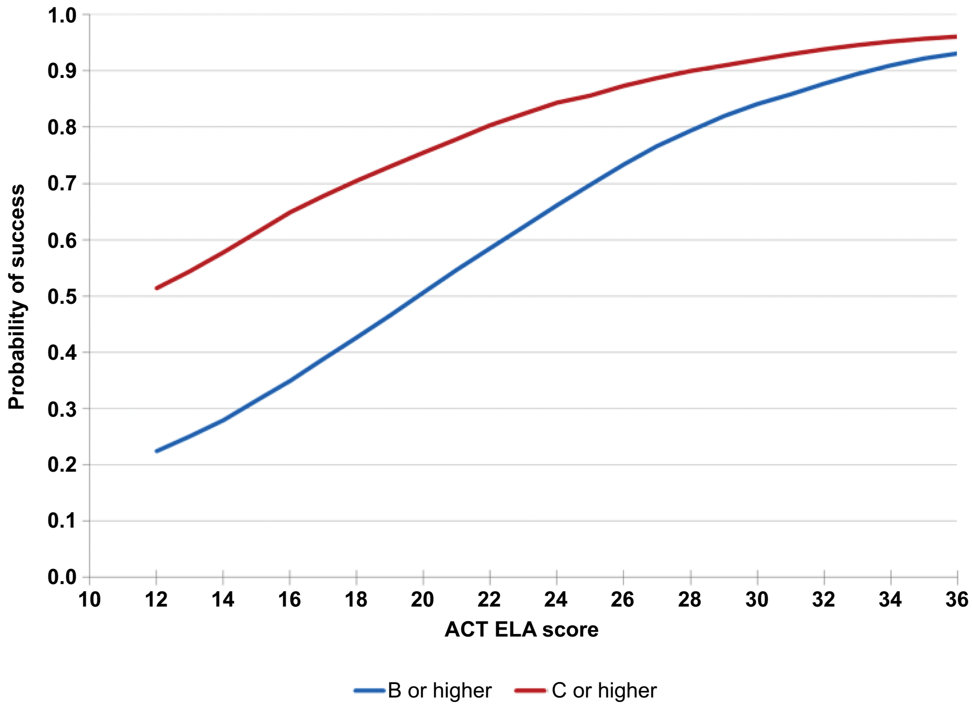


Figure 8.2 The typical probability of success in ELA-related courses by the ACT ELA score. The English related course was English Composition I. The social science courses included American History, Other History, Psychology, Sociology, Political Science, and Economics.

Summary. The ACT ELA Readiness Benchmark can be used to help gauge overall student readiness for ELA-related coursework. In the academic year of 2015–16, 519,922 students (25%) from the 2016 ACT-tested high school graduating class took the current ACT writing test, and so they had an official ACT ELA score. Of these students, 61% met the ACT ELA Benchmark of 20. Providing ELA readiness information based on students' English, reading, and writing skills to prospective students may help facilitate the transition to college by raising their awareness of the literacy skills required to meet the demands of the array of ELA-related courses they will face in college. Such feedback can send a signal to students as to the level of readiness needed to avoid having to take remedial course work in English and reading that can impede students' progress toward earning a college degree.

A limitation of the Radunzel et al. study (2017) was that its preliminary benchmark was based on estimated ELA scores using concorded ACT writing scores. There are plans to reevaluate the ELA Benchmark once sufficient college course-transcript data become available for students who took the current ACT writing test. That data set will include freshman cohorts of 2016 and later.

8.3.5 Intended Uses of the Benchmarks for Students, Schools, Districts, and States

ACT scores give students an indication of how likely they are to succeed in college-level courses. The results let students know if they have developed or are developing the foundation for the skills they will need by the time they finish high school.

In 2014, ACT launched ACT® Aspire, a test battery that measures students' mastery of English, mathematics, reading, and science in Grades 3 through 10. Readiness Benchmarks have been developed for ACT Aspire that indicate whether students are on target to meet the ACT College Readiness Benchmarks in Grade 11, allowing for the articulation of what students need to know and be able to do at key transition points along the K-Career continuum. Each ACT Aspire subject test has its own grade-level specific ACT Readiness Benchmarks. Students at or above the Benchmark are on target to meet the corresponding ACT College Readiness Benchmark in Grade 11, assuming that these students will continue to work hard and take challenging courses throughout high school. For more details about the development of the ACT Readiness Benchmarks used with ACT Aspire, see the ACT Aspire Technical Manual (ACT, 2016b).

Researchers and policymakers can use the Benchmarks to monitor the educational progress of schools, districts, and states. Middle and high school personnel can use the Benchmarks for ACT Aspire as a means of evaluating students' early progress toward college readiness so that timely interventions can be implemented when necessary and well before students approach high school graduation, or as an educational counseling or career-planning tool. Such information helps students and teachers know if a student is on track for college and career readiness.

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

The performance levels on the ACT tests 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 tools 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 needed to improve the likelihood of achieving college success.

ACT College and Career Readiness Standards—English

	Production of Writing		Knowledge of Language
	Topic Development in Terms of Purpose and Focus (TOD)	Organization, Unity, and Cohesion (ORG)	Knowledge of Language (KLA)
13–15	**201.** Delete material because it is obviously irrelevant in terms of the topic of the essay	**201.** Determine the need for transition words or phrases to establish time relationships in simple narrative essays (e.g., *then*, *this time*)	**201.** Revise vague, clumsy, and confusing writing that creates obvious logic problems
16–19	**301.** Delete material because it is obviously irrelevant in terms of the focus of the essay **302.** Identify the purpose of a word or phrase when the purpose is simple (e.g., identifying a person, defining a basic term, using common descriptive adjectives) **303.** Determine whether a simple essay has met a straightforward goal	**301.** Determine the most logical place for a sentence in a paragraph **302.** Provide a simple conclusion to a paragraph or essay (e.g., expressing one of the essay's main ideas)	**301.** Delete obviously redundant and wordy material **302.** Revise expressions that deviate markedly from the style and tone of the essay
20–23	**401.** Determine relevance of material in terms of the focus of the essay **402.** Identify the purpose of a word or phrase when the purpose is straightforward (e.g., describing a person, giving examples) **403.** Use a word, phrase, or sentence to accomplish a straightforward purpose (e.g., conveying a feeling or attitude)	**401.** Determine the need for transition words or phrases to establish straightforward logical relationships (e.g., *first*, *afterward*, *in response*) **402.** Determine the most logical place for a sentence in a straightforward essay **403.** Provide an introduction to a straightforward paragraph **404.** Provide a straightforward conclusion to a paragraph or essay (e.g., summarizing an essay's main idea or ideas) **405.** Rearrange the sentences in a straightforward paragraph for the sake of logic	**401.** Delete redundant and wordy material when the problem is contained within a single phrase (e.g., "alarmingly startled," "started by reaching the point of beginning") **402.** Revise expressions that deviate from the style and tone of the essay **403.** Determine the need for conjunctions to create straightforward logical links between clauses **404.** Use the word or phrase most appropriate in terms of the content of the sentence when the vocabulary is relatively common

ACT College and Career Readiness Standards—English—continued

	Production of Writing		Knowledge of Language
	Topic Development in Terms of Purpose and Focus (TOD)	Organization, Unity, and Cohesion (ORG)	Knowledge of Language (KLA)
24–27	**501.** Determine relevance of material in terms of the focus of the paragraph **502.** Identify the purpose of a word, phrase, or sentence when the purpose is fairly straightforward (e.g., identifying traits, giving reasons, explaining motivations) **503.** Determine whether an essay has met a specified goal **504.** Use a word, phrase, or sentence to accomplish a fairly straightforward purpose (e.g., sharpening an essay's focus, illustrating a given statement)	**501.** Determine the need for transition words or phrases to establish subtle logical relationships within and between sentences (e.g., *therefore*, *however*, *in addition*) **502.** Provide a fairly straightforward introduction or conclusion to or transition within a paragraph or essay (e.g., supporting or emphasizing an essay's main idea) **503.** Rearrange the sentences in a fairly straightforward paragraph for the sake of logic **504.** Determine the best place to divide a paragraph to meet a particular rhetorical goal **505.** Rearrange the paragraphs in an essay for the sake of logic	**501.** Revise vague, clumsy, and confusing writing **502.** Delete redundant and wordy material when the meaning of the entire sentence must be considered **503.** Revise expressions that deviate in subtle ways from the style and tone of the essay **504.** Determine the need for conjunctions to create logical links between clauses **505.** Use the word or phrase most appropriate in terms of the content of the sentence when the vocabulary is uncommon
28–32	**601.** Determine relevance when considering material that is plausible but potentially irrelevant at a given point in the essay **602.** Identify the purpose of a word, phrase, or sentence when the purpose is subtle (e.g., supporting a later point, establishing tone) or when the best decision is to delete the text in question **603.** Use a word, phrase, or sentence to accomplish a subtle purpose (e.g., adding emphasis or supporting detail, expressing meaning through connotation)	**601.** Determine the need for transition words or phrases to establish subtle logical relationships within and between paragraphs **602.** Determine the most logical place for a sentence in a fairly complex essay **603.** Provide a subtle introduction or conclusion to or transition within a paragraph or essay (e.g., echoing an essay's theme or restating the main argument) **604.** Rearrange the sentences in a fairly complex paragraph for the sake of logic and coherence	**601.** Revise vague, clumsy, and confusing writing involving sophisticated language **602.** Delete redundant and wordy material that involves fairly sophisticated language (e.g., "the outlook of an aesthetic viewpoint") or that sounds acceptable as conversational English **603.** Determine the need for conjunctions to create subtle logical links between clauses **604.** Use the word or phrase most appropriate in terms of the content of the sentence when the vocabulary is fairly sophisticated

ACT College and Career Readiness Standards—English—continued

	Production of Writing		Knowledge of Language
	Topic Development in Terms of Purpose and Focus (TOD)	Organization, Unity, and Cohesion (ORG)	Knowledge of Language (KLA)
33–36	**701.** Identify the purpose of a word, phrase, or sentence when the purpose is complex (e.g., anticipating a reader's need for background information) or requires a thorough understanding of the paragraph and essay **702.** Determine whether a complex essay has met a specified goal **703.** Use a word, phrase, or sentence to accomplish a complex purpose, often in terms of the focus of the essay	**701.** Determine the need for transition words or phrases, basing decisions on a thorough understanding of the paragraph and essay **702.** Provide a sophisticated introduction or conclusion to or transition within a paragraph or essay, basing decisions on a thorough understanding of the paragraph and essay (e.g., linking the conclusion to one of the essay's main images)	**701.** Delete redundant and wordy material that involves sophisticated language or complex concepts or where the material is redundant in terms of the paragraph or essay as a whole **702.** Use the word or phrase most appropriate in terms of the content of the sentence when the vocabulary is sophisticated

ACT College and Career Readiness Standards—English—continued

	Conventions of Standard English Grammar, Usage, and Punctuation		
	Sentence Structure and Formation (SST)	Usage Conventions (USG)	Punctuation Conventions (PUN)
13–15	**201.** Determine the need for punctuation or conjunctions to join simple clauses **202.** Recognize and correct inappropriate shifts in verb tense between simple clauses in a sentence or between simple adjoining sentences	**201.** Form the past tense and past participle of irregular but commonly used verbs **202.** Form comparative and superlative adjectives	**201.** Delete commas that create basic sense problems (e.g., between verb and direct object)
16–19	**301.** Determine the need for punctuation or conjunctions to correct awkward-sounding fragments and fused sentences as well as obviously faulty subordination and coordination of clauses **302.** Recognize and correct inappropriate shifts in verb tense and voice when the meaning of the entire sentence must be considered	**301.** Determine whether an adjective form or an adverb form is called for in a given situation **302.** Ensure straightforward subject-verb agreement **303.** Ensure straightforward pronoun-antecedent agreement **304.** Use idiomatically appropriate prepositions in simple contexts **305.** Use the appropriate word in frequently confused pairs (e.g., *there* and *their*, *past* and *passed*, *led* and *lead*)	**301.** Delete commas that markedly disturb sentence flow (e.g., between modifier and modified element) **302.** Use appropriate punctuation in straightforward situations (e.g., simple items in a series)
20–23	**401.** Recognize and correct marked disturbances in sentence structure (e.g., faulty placement of adjectives, participial phrase fragments, missing or incorrect relative pronouns, dangling or misplaced modifiers, lack of parallelism within a simple series of verbs)	**401.** Use the correct comparative or superlative adjective or adverb form depending on context (e.g., "He is the oldest of my three brothers") **402.** Ensure subject-verb agreement when there is some text between the subject and verb **403.** Use idiomatically appropriate prepositions, especially in combination with verbs (e.g., *long for*, *appeal to*) **404.** Recognize and correct expressions that deviate from idiomatic English	**401.** Delete commas when an incorrect understanding of the sentence suggests a pause that should be punctuated (e.g., between verb and direct object clause) **402.** Delete apostrophes used incorrectly to form plural nouns **403.** Use commas to avoid obvious ambiguity (e.g., to set off a long introductory element from the rest of the sentence when a misreading is possible) **404.** Use commas to set off simple parenthetical elements

ACT College and Career Readiness Standards—English—continued

	Conventions of Standard English Grammar, Usage, and Punctuation		
	Sentence Structure and Formation (SST)	Usage Conventions (USG)	Punctuation Conventions (PUN)
24–27	**501.** Recognize and correct disturbances in sentence structure (e.g., faulty placement of phrases, faulty coordination and subordination of clauses, lack of parallelism within a simple series of phrases) **502.** Maintain consistent and logical verb tense and pronoun person on the basis of the preceding clause or sentence	**501.** Form simple and compound verb tenses, both regular and irregular, including forming verbs by using *have* rather than *of* (e.g., *would have gone*, not *would of gone*) **502.** Ensure pronoun-antecedent agreement when the pronoun and antecedent occur in separate clauses or sentences **503.** Recognize and correct vague and ambiguous pronouns	**501.** Delete commas in long or involved sentences when an incorrect understanding of the sentence suggests a pause that should be punctuated (e.g., between the elements of a compound subject or compound verb joined by *and*) **502.** Recognize and correct inappropriate uses of colons and semicolons **503.** Use punctuation to set off complex parenthetical elements **504.** Use apostrophes to form simple possessive nouns
28–32	**601.** Recognize and correct subtle disturbances in sentence structure (e.g., danglers where the intended meaning is clear but the sentence is ungrammatical, faulty subordination and coordination of clauses in long or involved sentences) **602.** Maintain consistent and logical verb tense and voice and pronoun person on the basis of the paragraph or essay as a whole	**601.** Ensure subject-verb agreement in some challenging situations (e.g., when the subject-verb order is inverted or when the subject is an indefinite pronoun) **602.** Correctly use reflexive pronouns, the possessive pronouns *its* and *your*, and the relative pronouns *who* and *whom* **603.** Use the appropriate word in less-common confused pairs (e.g., *allude* and *elude*)	**601.** Use commas to avoid ambiguity when the syntax or language is sophisticated (e.g., to set off a complex series of items) **602.** Use punctuation to set off a nonessential/nonrestrictive appositive or clause **603.** Use apostrophes to form possessives, including irregular plural nouns **604.** Use a semicolon to link closely related independent clauses

ACT College and Career Readiness Standards—English—continued

	Conventions of Standard English Grammar, Usage, and Punctuation		
	Sentence Structure and Formation (SST)	Usage Conventions (USG)	Punctuation Conventions (PUN)
33–36	**701.** Recognize and correct very subtle disturbances in sentence structure (e.g., weak conjunctions between independent clauses, run-ons that would be acceptable in conversational English, lack of parallelism within a complex series of phrases or clauses)	**701.** Ensure subject-verb agreement when a phrase or clause between the subject and verb suggests a different number for the verb **702.** Use idiomatically and contextually appropriate prepositions in combination with verbs in situations involving sophisticated language or complex concepts	**701.** Delete punctuation around essential/restrictive appositives or clauses **702.** Use a colon to introduce an example or an elaboration

ACT College and Career Readiness Standards—Mathematics

	Topics in the flow to Number and Quantity	Topics in the flow to Algebra	Topics in the flow to Functions
13–15	**N 201.** Perform one-operation computation with whole numbers and decimals **N 202.** Recognize equivalent fractions and fractions in lowest terms **N 203.** Locate positive rational numbers (expressed as whole numbers, fractions, decimals, and mixed numbers) on the number line	**AF 201.** Solve problems in one or two steps using whole numbers and using decimals in the context of money **A 201.** Exhibit knowledge of basic expressions (e.g., identify an expression for a total as $b + g$) **A 202.** Solve equations in the form $x + a = b$, where a and b are whole numbers or decimals	**F 201.** Extend a given pattern by a few terms for patterns that have a constant increase or decrease between terms
16–19	**N 301.** Recognize one-digit factors of a number **N 302.** Identify a digit's place value **N 303.** Locate rational numbers on the number line *Note: A matrix as a representation of data is treated here as a basic table.*	**AF 301.** Solve routine one-step arithmetic problems using positive rational numbers, such as single-step percent **AF 302.** Solve some routine two-step arithmetic problems **AF 303.** Relate a graph to a situation described qualitatively in terms of familiar properties such as before and after, increasing and decreasing, higher and lower **AF 304.** Apply a definition of an operation for whole numbers (e.g., $a \square b = 3a - b$) **A 301.** Substitute whole numbers for unknown quantities to evaluate expressions **A 302.** Solve one-step equations to get integer or decimal answers **A 303.** Combine like terms (e.g., $2x + 5x$)	**F 301.** Extend a given pattern by a few terms for patterns that have a constant factor between terms

Note: AF standards span the Algebra and Functions columns.

ACT College and Career Readiness Standards—Mathematics—continued

<table>
<tr><th></th><th>Topics in the flow to
Number and Quantity</th><th>Topics in the flow to
Algebra</th><th>Topics in the flow to
Functions</th></tr>
<tr><td rowspan="2">20–23</td><td rowspan="2">N 401. Exhibit knowledge of elementary number concepts such as rounding, the ordering of decimals, pattern identification, primes, and greatest common factor
N 402. Write positive powers of 10 by using exponents
N 403. Comprehend the concept of length on the number line, and find the distance between two points
N 404. Understand absolute value in terms of distance
N 405. Find the distance in the coordinate plane between two points with the same x-coordinate or y-coordinate
N 406. Add two matrices that have whole number entries</td><td colspan="2">AF 401. Solve routine two-step or three-step arithmetic problems involving concepts such as rate and proportion, tax added, percentage off, and estimating by using a given average value in place of actual values
AF 402. Perform straightforward word-to-symbol translations
AF 403. Relate a graph to a situation described in terms of a starting value and an additional amount per unit (e.g., unit cost, weekly growth)</td></tr>
<tr><td>A 401. Evaluate algebraic expressions by substituting integers for unknown quantities
A 402. Add and subtract simple algebraic expressions
A 403. Solve routine first-degree equations
A 404. Multiply two binomials
A 405. Match simple inequalities with their graphs on the number line (e.g., $x \ge -\frac{3}{5}$)
A 406. Exhibit knowledge of slope</td><td>F 401. Evaluate linear and quadratic functions, expressed in function notation, at integer values</td></tr>
</table>

ACT College and Career Readiness Standards—Mathematics—continued

	Topics in the flow to Number and Quantity	Topics in the flow to Algebra	Topics in the flow to Functions
24–27	**N 501.** Order fractions **N 502.** Find and use the least common multiple **N 503.** Work with numerical factors **N 504.** Exhibit some knowledge of the complex numbers **N 505.** Add and subtract matrices that have integer entries	**AF 501.** Solve multistep arithmetic problems that involve planning or converting common derived units of measure (e.g., feet per second to miles per hour) **AF 502.** Build functions and write expressions, equations, or inequalities with a single variable for common pre-algebra settings (e.g., rate and distance problems and problems that can be solved by using proportions) **AF 503.** Match linear equations with their graphs in the coordinate plane	
		A 501. Recognize that when numerical quantities are reported in real-world contexts, the numbers are often rounded **A 502.** Solve real-world problems by using first-degree equations **A 503.** Solve first-degree inequalities when the method does not involve reversing the inequality sign **A 504.** Match compound inequalities with their graphs on the number line (e.g., $-10.5 < x \le 20.3$) **A 505.** Add, subtract, and multiply polynomials **A 506.** Identify solutions to simple quadratic equations **A 507.** Solve quadratic equations in the form $(x + a)(x + b) = 0$, where a and b are numbers or variables **A 508.** Factor simple quadratics (e.g., the difference of squares and perfect square trinomials) **A 509.** Work with squares and square roots of numbers **A 510.** Work with cubes and cube roots of numbers **A 511.** Work with scientific notation **A 512.** Work problems involving positive integer exponents **A 513.** Determine when an expression is undefined **A 514.** Determine the slope of a line from an equation	**F 501.** Evaluate polynomial functions, expressed in function notation, at integer values **F 502.** Find the next term in a sequence described recursively **F 503.** Build functions and use quantitative information to identify graphs for relations that are proportional or linear **F 504.** Attend to the difference between a function modeling a situation and the reality of the situation **F 505.** Understand the concept of a function as having a well-defined output value at each valid input value **F 506.** Understand the concept of domain and range in terms of valid input and output, and in terms of function graphs **F 507.** Interpret statements that use function notation in terms of their context **F 508.** Find the domain of polynomial functions and rational functions **F 509.** Find the range of polynomial functions **F 510.** Find where a rational function's graph has a vertical asymptote **F 511.** Use function notation for simple functions of two variables

ACT College and Career Readiness Standards—Mathematics—continued

	Topics in the flow to Number and Quantity	Topics in the flow to Algebra	Topics in the flow to Functions
28–32	**N 601.** Apply number properties involving prime factorization **N 602.** Apply number properties involving even/odd numbers and factors/multiples **N 603.** Apply number properties involving positive/negative numbers **N 604.** Apply the facts that π is irrational and that the square root of an integer is rational only if that integer is a perfect square **N 605.** Apply properties of rational exponents **N 606.** Multiply two complex numbers **N 607.** Use relations involving addition, subtraction, and scalar multiplication of vectors and of matrices	**AF 601.** Solve word problems containing several rates, proportions, or percentages **AF 602.** Build functions and write expressions, equations, and inequalities for common algebra settings (e.g., distance to a point on a curve and profit for variable cost and demand) **AF 603.** Interpret and use information from graphs in the coordinate plane **AF 604.** Given an equation or function, find an equation or function whose graph is a translation by a specified amount up or down **A 601.** Manipulate expressions and equations **A 602.** Solve linear inequalities when the method involves reversing the inequality sign **A 603.** Match linear inequalities with their graphs on the number line **A 604.** Solve systems of two linear equations **A 605.** Solve quadratic equations **A 606.** Solve absolute value equations	(AF 601–AF 604 span Algebra and Functions) **F 601.** Relate a graph to a situation described qualitatively in terms of faster change or slower change **F 602.** Build functions for relations that are inversely proportional **F 603.** Find a recursive expression for the general term in a sequence described recursively **F 604.** Evaluate composite functions at integer values

ACT College and Career Readiness Standards—Mathematics—continued

	Topics in the flow to Number and Quantity	Topics in the flow to Algebra	Topics in the flow to Functions
33–36	**N 701.** Analyze and draw conclusions based on number concepts **N 702.** Apply properties of rational numbers and the rational number system **N 703.** Apply properties of real numbers and the real number system, including properties of irrational numbers **N 704.** Apply properties of complex numbers and the complex number system **N 705.** Multiply matrices **N 706.** Apply properties of matrices and properties of matrices as a number system	**AF 701.** Solve complex arithmetic problems involving percent of increase or decrease or requiring integration of several concepts (e.g., using several ratios, comparing percentages, or comparing averages) **AF 702.** Build functions and write expressions, equations, and inequalities when the process requires planning and/or strategic manipulation **AF 703.** Analyze and draw conclusions based on properties of algebra and/or functions **AF 704.** Analyze and draw conclusions based on information from graphs in the coordinate plane **AF 705.** Identify characteristics of graphs based on a set of conditions or on a general equation such as $y = ax^2 + c$ **AF 706.** Given an equation or function, find an equation or function whose graph is a translation by specified amounts in the horizontal and vertical directions	
		A 701. Solve simple absolute value inequalities **A 702.** Match simple quadratic inequalities with their graphs on the number line **A 703.** Apply the remainder theorem for polynomials, that $P(a)$ is the remainder when $P(x)$ is divided by $(x - a)$	**F 701.** Compare actual values and the values of a modeling function to judge model fit and compare models **F 702.** Build functions for relations that are exponential **F 703.** Exhibit knowledge of geometric sequences **F 704.** Exhibit knowledge of unit circle trigonometry **F 705.** Match graphs of basic trigonometric functions with their equations **F 706.** Use trigonometric concepts and basic identities to solve problems **F 707.** Exhibit knowledge of logarithms **F 708.** Write an expression for the composite of two simple functions

ACT College and Career Readiness Standards—Mathematics

	Topics in the flow to Geometry	Topics in the flow to Statistics and Probability
13–15	**G 201.** Estimate the length of a line segment based on other lengths in a geometric figure **G 202.** Calculate the length of a line segment based on the lengths of other line segments that go in the same direction (e.g., overlapping line segments and parallel sides of polygons with only right angles) **G 203.** Perform common conversions of money and of length, weight, mass, and time within a measurement system (e.g., dollars to dimes, inches to feet, and hours to minutes)	**S 201.** Calculate the average of a list of positive whole numbers **S 202.** Extract one relevant number from a basic table or chart, and use it in a single computation
16–19	**G 301.** Exhibit some knowledge of the angles associated with parallel lines **G 302.** Compute the perimeter of polygons when all side lengths are given **G 303.** Compute the area of rectangles when whole number dimensions are given **G 304.** Locate points in the first quadrant	**S 301.** Calculate the average of a list of numbers **S 302.** Calculate the average given the number of data values and the sum of the data values **S 303.** Read basic tables and charts **S 304.** Extract relevant data from a basic table or chart and use the data in a computation **S 305.** Use the relationship between the probability of an event and the probability of its complement
20–23	**G 401.** Use properties of parallel lines to find the measure of an angle **G 402.** Exhibit knowledge of basic angle properties and special sums of angle measures (e.g., 90°, 180°, and 360°) **G 403.** Compute the area and perimeter of triangles and rectangles in simple problems **G 404.** Find the length of the hypotenuse of a right triangle when only very simple computation is involved (e.g., 3-4-5 and 6-8-10 triangles) **G 405.** Use geometric formulas when all necessary information is given **G 406.** Locate points in the coordinate plane **G 407.** Translate points up, down, left, and right in the coordinate plane	**S 401.** Calculate the missing data value given the average and all data values but one **S 402.** Translate from one representation of data to another (e.g., a bar graph to a circle graph) **S 403.** Determine the probability of a simple event **S 404.** Describe events as combinations of other events (e.g., using *and*, *or*, and *not*) **S 405.** Exhibit knowledge of simple counting techniques

ACT College and Career Readiness Standards—Mathematics—continued

	Topics in the flow to Geometry	Topics in the flow to Statistics and Probability
24–27	**G 501.** Use several angle properties to find an unknown angle measure **G 502.** Count the number of lines of symmetry of a geometric figure **G 503.** Use symmetry of isosceles triangles to find unknown side lengths or angle measures **G 504.** Recognize that real-world measurements are typically imprecise and that an appropriate level of precision is related to the measuring device and procedure **G 505.** Compute the perimeter of simple composite geometric figures with unknown side lengths **G 506.** Compute the area of triangles and rectangles when one or more additional simple steps are required **G 507.** Compute the area and circumference of circles after identifying necessary information **G 508.** Given the length of two sides of a right triangle, find the third when the lengths are Pythagorean triples **G 509.** Express the sine, cosine, and tangent of an angle in a right triangle as a ratio of given side lengths **G 510.** Determine the slope of a line from points or a graph **G 511.** Find the midpoint of a line segment **G 512.** Find the coordinates of a point rotated 180° around a given center point	**S 501.** Calculate the average given the frequency counts of all the data values **S 502.** Manipulate data from tables and charts **S 503.** Compute straightforward probabilities for common situations **S 504.** Use Venn diagrams in counting **S 505.** Recognize that when data summaries are reported in the real world, results are often rounded and must be interpreted as having appropriate precision **S 506.** Recognize that when a statistical model is used, model values typically differ from actual values

ACT College and Career Readiness Standards—Mathematics—continued

	Topics in the flow to Geometry	Topics in the flow to Statistics and Probability
28–32	**G 601.** Use relationships involving area, perimeter, and volume of geometric figures to compute another measure (e.g., surface area for a cube of a given volume and simple geometric probability) **G 602.** Use the Pythagorean theorem **G 603.** Apply properties of 30°-60°-90°, 45°-45°-90°, similar, and congruent triangles **G 604.** Apply basic trigonometric ratios to solve right-triangle problems **G 605.** Use the distance formula **G 606.** Use properties of parallel and perpendicular lines to determine an equation of a line or coordinates of a point **G 607.** Find the coordinates of a point reflected across a vertical or horizontal line or across $y = x$ **G 608.** Find the coordinates of a point rotated 90° about the origin **G 609.** Recognize special characteristics of parabolas and circles (e.g., the vertex of a parabola and the center or radius of a circle)	**S 601.** Calculate or use a weighted average **S 602.** Interpret and use information from tables and charts, including two-way frequency tables **S 603.** Apply counting techniques **S 604.** Compute a probability when the event and/or sample space are not given or obvious **S 605.** Recognize the concepts of conditional and joint probability expressed in real-world contexts **S 606.** Recognize the concept of independence expressed in real-world contexts
33–36	**G 701.** Use relationships among angles, arcs, and distances in a circle **G 702.** Compute the area of composite geometric figures when planning and/or visualization is required **G 703.** Use scale factors to determine the magnitude of a size change **G 704.** Analyze and draw conclusions based on a set of conditions **G 705.** Solve multistep geometry problems that involve integrating concepts, planning, and/or visualization	**S 701.** Distinguish between mean, median, and mode for a list of numbers **S 702.** Analyze and draw conclusions based on information from tables and charts, including two-way frequency tables **S 703.** Understand the role of randomization in surveys, experiments, and observational studies **S 704.** Exhibit knowledge of conditional and joint probability **S 705.** Recognize that part of the power of statistical modeling comes from looking at regularity in the differences between actual values and model values

ACT College and Career Readiness Standards—Reading

	Key Ideas and Details			
	Close Reading (CLR)	Central Ideas, Themes, and Summaries (IDT)	Relationships (REL)	Word Meanings and Word Choice (WME)
13–15	**201.** Locate basic facts (e.g., names, dates, events) clearly stated in a passage **202.** Draw simple logical conclusions about the main characters in somewhat challenging literary narratives	**201.** Identify the topic of passages and distinguish the topic from the central idea or theme	**201.** Determine when (e.g., first, last, before, after) an event occurs in somewhat challenging passages **202.** Identify simple cause-effect relationships within a single sentence in a passage	**201.** Understand the implication of a familiar word or phrase and of simple descriptive language
16–19	**301.** Locate simple details at the sentence and paragraph level in somewhat challenging passages **302.** Draw simple logical conclusions in somewhat challenging passages	**301.** Identify a clear central idea in straightforward paragraphs in somewhat challenging literary narratives	**301.** Identify clear comparative relationships between main characters in somewhat challenging literary narratives **302.** Identify simple cause-effect relationships within a single paragraph in somewhat challenging literary narratives	**301.** Analyze how the choice of a specific word or phrase shapes meaning or tone in somewhat challenging passages when the effect is simple **302.** Interpret basic figurative language as it is used in a passage
20–23	**401.** Locate important details in somewhat challenging passages **402.** Draw logical conclusions in somewhat challenging passages **403.** Draw simple logical conclusions in more challenging passages **404.** Paraphrase some statements as they are used in somewhat challenging passages	**401.** Infer a central idea in straightforward paragraphs in somewhat challenging literary narratives **402.** Identify a clear central idea or theme in somewhat challenging passages or their paragraphs **403.** Summarize key supporting ideas and details in somewhat challenging passages	**401.** Order simple sequences of events in somewhat challenging literary narratives **402.** Identify clear comparative relationships in somewhat challenging passages **403.** Identify clear cause-effect relationships in somewhat challenging passages	**401.** Analyze how the choice of a specific word or phrase shapes meaning or tone in somewhat challenging passages **402.** Interpret most words and phrases as they are used in somewhat challenging passages, including determining technical, connotative, and figurative meanings

ACT College and Career Readiness Standards—Reading—continued

	Key Ideas and Details			
	Close Reading (CLR)	Central Ideas, Themes, and Summaries (IDT)	Relationships (REL)	Word Meanings and Word Choice (WME)
24–27	**501.** Locate and interpret minor or subtly stated details in somewhat challenging passages **502.** Locate important details in more challenging passages **503.** Draw subtle logical conclusions in somewhat challenging passages **504.** Draw logical conclusions in more challenging passages **505.** Paraphrase virtually any statement as it is used in somewhat challenging passages **506.** Paraphrase some statements as they are used in more challenging passages	**501.** Infer a central idea or theme in somewhat challenging passages or their paragraphs **502.** Identify a clear central idea or theme in more challenging passages or their paragraphs **503.** Summarize key supporting ideas and details in more challenging passages	**501.** Order sequences of events in somewhat challenging passages **502.** Understand implied or subtly stated comparative relationships in somewhat challenging passages **503.** Identify clear comparative relationships in more challenging passages **504.** Understand implied or subtly stated cause-effect relationships in somewhat challenging passages **505.** Identify clear cause-effect relationships in more challenging passages	**501.** Analyze how the choice of a specific word or phrase shapes meaning or tone in somewhat challenging passages when the effect is subtle **502.** Analyze how the choice of a specific word or phrase shapes meaning or tone in more challenging passages **503.** Interpret virtually any word or phrase as it is used in somewhat challenging passages, including determining technical, connotative, and figurative meanings **504.** Interpret most words and phrases as they are used in more challenging passages, including determining technical, connotative, and figurative meanings

ACT College and Career Readiness Standards—Reading—continued

	Key Ideas and Details			
	Close Reading (CLR)	Central Ideas, Themes, and Summaries (IDT)	Relationships (REL)	Word Meanings and Word Choice (WME)
28–32	**601.** Locate and interpret minor or subtly stated details in more challenging passages **602.** Locate important details in complex passages **603.** Draw subtle logical conclusions in more challenging passages **604.** Draw simple logical conclusions in complex passages **605.** Paraphrase virtually any statement as it is used in more challenging passages	**601.** Infer a central idea or theme in more challenging passages or their paragraphs **602.** Summarize key supporting ideas and details in complex passages	**601.** Order sequences of events in more challenging passages **602.** Understand implied or subtly stated comparative relationships in more challenging passages **603.** Identify clear comparative relationships in complex passages **604.** Understand implied or subtly stated cause-effect relationships in more challenging passages **605.** Identify clear cause-effect relationships in complex passages	**601.** Analyze how the choice of a specific word or phrase shapes meaning or tone in complex passages **602.** Interpret virtually any word or phrase as it is used in more challenging passages, including determining technical, connotative, and figurative meanings **603.** Interpret words and phrases in a passage that makes consistent use of figurative, general academic, domain-specific, or otherwise difficult language
33–36	**701.** Locate and interpret minor or subtly stated details in complex passages **702.** Locate important details in highly complex passages **703.** Draw logical conclusions in complex passages **704.** Draw simple logical conclusions in highly complex passages **705.** Draw complex or subtle logical conclusions, often by synthesizing information from different portions of the passage **706.** Paraphrase statements as they are used in complex passages	**701.** Identify or infer a central idea or theme in complex passages or their paragraphs **702.** Summarize key supporting ideas and details in highly complex passages	**701.** Order sequences of events in complex passages **702.** Understand implied or subtly stated comparative relationships in complex passages **703.** Identify clear comparative relationships in highly complex passages **704.** Understand implied or subtly stated cause-effect relationships in complex passages **705.** Identify clear cause-effect relationships in highly complex passages	**701.** Analyze how the choice of a specific word or phrase shapes meaning or tone in passages when the effect is subtle or complex **702.** Interpret words and phrases as they are used in complex passages, including determining technical, connotative, and figurative meanings **703.** Interpret words and phrases in a passage that makes extensive use of figurative, general academic, domain-specific, or otherwise difficult language

ACT College and Career Readiness Standards—Reading—continued

	Craft and Structure		Integration of Knowledge and Ideas	
	Text Structure (TST)	Purpose and Point of View (PPV)	Arguments (ARG)	Multiple Texts (SYN)
13–15	**201.** Analyze how one or more sentences in passages relate to the whole passage when the function is stated or clearly indicated	**201.** Recognize a clear intent of an author or narrator in somewhat challenging literary narratives	**201.** Analyze how one or more sentences in passages offer reasons for or support a claim when the relationship is clearly indicated	**201.** Make simple comparisons between two passages
16–19	**301.** Analyze how one or more sentences in somewhat challenging passages relate to the whole passage when the function is simple **302.** Identify a clear function of straightforward paragraphs in somewhat challenging literary narratives	**301.** Recognize a clear intent of an author or narrator in somewhat challenging passages	**301.** Analyze how one or more sentences in somewhat challenging passages offer reasons for or support a claim when the relationship is simple	**301.** Make straightforward comparisons between two passages
20–23	**401.** Analyze how one or more sentences in somewhat challenging passages relate to the whole passage **402.** Infer the function of straightforward paragraphs in somewhat challenging literary narratives **403.** Identify a clear function of paragraphs in somewhat challenging passages **404.** Analyze the overall structure of somewhat challenging passages	**401.** Identify a clear purpose of somewhat challenging passages and how that purpose shapes content and style **402.** Understand point of view in somewhat challenging passages	**401.** Analyze how one or more sentences in somewhat challenging passages offer reasons for or support a claim **402.** Identify a clear central claim in somewhat challenging passages	**401.** Draw logical conclusions using information from two literary narratives

ACT College and Career Readiness Standards—Reading—continued

	Craft and Structure		Integration of Knowledge and Ideas	
	Text Structure (TST)	Purpose and Point of View (PPV)	Arguments (ARG)	Multiple Texts (SYN)
24–27	**501.** Analyze how one or more sentences in somewhat challenging passages relate to the whole passage when the function is subtle **502.** Analyze how one or more sentences in more challenging passages relate to the whole passage **503.** Infer the function of paragraphs in somewhat challenging passages **504.** Identify a clear function of paragraphs in more challenging passages **505.** Analyze the overall structure of more challenging passages	**501.** Infer a purpose in somewhat challenging passages and how that purpose shapes content and style **502.** Identify a clear purpose of more challenging passages and how that purpose shapes content and style **503.** Understand point of view in more challenging passages	**501.** Analyze how one or more sentences in more challenging passages offer reasons for or support a claim **502.** Infer a central claim in somewhat challenging passages **503.** Identify a clear central claim in more challenging passages	**501.** Draw logical conclusions using information from two informational texts
28–32	**601.** Analyze how one or more sentences in complex passages relate to the whole passage **602.** Infer the function of paragraphs in more challenging passages **603.** Analyze the overall structure of complex passages	**601.** Infer a purpose in more challenging passages and how that purpose shapes content and style **602.** Understand point of view in complex passages	**601.** Analyze how one or more sentences in complex passages offer reasons for or support a claim **602.** Infer a central claim in more challenging passages	**601.** Draw logical conclusions using information from multiple portions of two literary narratives

ACT College and Career Readiness Standards—Reading—continued

	Craft and Structure		Integration of Knowledge and Ideas	
	Text Structure (TST)	Purpose and Point of View (PPV)	Arguments (ARG)	Multiple Texts (SYN)
33–36	**701.** Analyze how one or more sentences in passages relate to the whole passage when the function is subtle or complex **702.** Identify or infer the function of paragraphs in complex passages **703.** Analyze the overall structure of highly complex passages	**701.** Identify or infer a purpose in complex passages and how that purpose shapes content and style **702.** Understand point of view in highly complex passages	**701.** Analyze how one or more sentences in passages offer reasons for or support a claim when the relationship is subtle or complex **702.** Identify or infer a central claim in complex passages **703.** Identify a clear central claim in highly complex passages	**701.** Draw logical conclusions using information from multiple portions of two informational texts

ACT College and Career Readiness Standards—Science

	Interpretation of Data (IOD)	Scientific Investigation (SIN)	Evaluation of Models, Inferences, and Experimental Results (EMI)
13–15	**201.** Select one piece of data from a simple data presentation (e.g., a simple food web diagram) **202.** Identify basic features of a table, graph, or diagram (e.g., units of measurement) **203.** Find basic information in text that describes a simple data presentation	**201.** Find basic information in text that describes a simple experiment **202.** Understand the tools and functions of tools used in a simple experiment	**201.** Find basic information in a model (conceptual)
16–19	**301.** Select two or more pieces of data from a simple data presentation **302.** Understand basic scientific terminology **303.** Find basic information in text that describes a complex data presentation **304.** Determine how the values of variables change as the value of another variable changes in a simple data presentation	**301.** Understand the methods used in a simple experiment **302.** Understand the tools and functions of tools used in a complex experiment **303.** Find basic information in text that describes a complex experiment	**301.** Identify implications in a model **302.** Determine which models present certain basic information
20–23	**401.** Select data from a complex data presentation (e.g., a phase diagram) **402.** Compare or combine data from a simple data presentation (e.g., order or sum data from a table) **403.** Translate information into a table, graph, or diagram **404.** Perform a simple interpolation or simple extrapolation using data in a table or graph	**401.** Understand a simple experimental design **402.** Understand the methods used in a complex experiment **403.** Identify a control in an experiment **404.** Identify similarities and differences between experiments **405.** Determine which experiments utilized a given tool, method, or aspect of design	**401.** Determine which simple hypothesis, prediction, or conclusion is, or is not, consistent with a data presentation, model, or piece of information in text **402.** Identify key assumptions in a model **403.** Determine which models imply certain information **404.** Identify similarities and differences between models

ACT College and Career Readiness Standards—Science—continued

	Interpretation of Data (IOD)	Scientific Investigation (SIN)	Evaluation of Models, Inferences, and Experimental Results (EMI)
24–27	**501.** Compare or combine data from two or more simple data presentations (e.g., categorize data from a table using a scale from another table) **502.** Compare or combine data from a complex data presentation **503.** Determine how the values of variables change as the value of another variable changes in a complex data presentation **504.** Determine and/or use a simple (e.g., linear) mathematical relationship that exists between data **505.** Analyze presented information when given new, simple information	**501.** Understand a complex experimental design **502.** Predict the results of an additional trial or measurement in an experiment **503.** Determine the experimental conditions that would produce specified results	**501.** Determine which simple hypothesis, prediction, or conclusion is, or is not, consistent with two or more data presentations, models, and/or pieces of information in text **502.** Determine whether presented information, or new information, supports or contradicts a simple hypothesis or conclusion, and why **503.** Identify the strengths and weaknesses of models **504.** Determine which models are supported or weakened by new information **505.** Determine which experimental results or models support or contradict a hypothesis, prediction, or conclusion
28–32	**601.** Compare or combine data from a simple data presentation with data from a complex data presentation **602.** Determine and/or use a complex (e.g., nonlinear) mathematical relationship that exists between data **603.** Perform a complex interpolation or complex extrapolation using data in a table or graph	**601.** Determine the hypothesis for an experiment **602.** Determine an alternate method for testing a hypothesis	**601.** Determine which complex hypothesis, prediction, or conclusion is, or is not, consistent with a data presentation, model, or piece of information in text **602.** Determine whether presented information, or new information, supports or weakens a model, and why **603.** Use new information to make a prediction based on a model

ACT College and Career Readiness Standards—Science—continued

	Interpretation of Data (IOD)	Scientific Investigation (SIN)	Evaluation of Models, Inferences, and Experimental Results (EMI)
33–36	**701.** Compare or combine data from two or more complex data presentations **702.** Analyze presented information when given new, complex information	**701.** Understand precision and accuracy issues **702.** Predict the effects of modifying the design or methods of an experiment **703.** Determine which additional trial or experiment could be performed to enhance or evaluate experimental results	**701.** Determine which complex hypothesis, prediction, or conclusion is, or is not, consistent with two or more data presentations, models, and/or pieces of information in text **702.** Determine whether presented information, or new information, supports or contradicts a complex hypothesis or conclusion, and why

ACT College and Career Readiness Standards for Science are measured in rich and authentic contexts based on science content that students encounter in science courses. This content includes:

Life Science/Biology	Physical Science/ Chemistry, Physics	Earth and Space Science
• Animal behavior • Animal development and growth • Body systems • Cell structure and processes • Ecology • Evolution • Genetics • Homeostasis • Life cycles • Molecular basis of heredity • Origin of life • Photosynthesis • Plant development, growth, structure • Populations • Taxonomy	• Atomic structure • Chemical bonding, equations, nomenclature, reactions • Electrical circuits • Elements, compounds, mixtures • Force and motions • Gravitation • Heat and work • Kinetic and potential energy • Magnetism • Momentum • The periodic table • Properties of solutions • Sound and light • States, classes, and properties of matter • Waves	• Earthquakes and volcanoes • Earth's atmosphere • Earth's resources • Fossils and geological time • Geochemical cycles • Groundwater • Lakes, rivers, oceans • Mass movements • Plate tectonics • Rocks, minerals • Solar system • Stars, galaxies, and the universe • Water cycle • Weather and climate • Weathering and erosion

ACT College and Career Readiness Standards—Writing

	Ideas and Analysis (IOD)	Development and Support (SIN)	Organization	Language Use and Conventions
11-12	Generate a nuanced, precise thesis that establishes a perspective on a contemporary issue Engage critically with other perspectives on the issue Establish and employ an insightful context for analysis Examine implications, complexities and tensions, and/or underlying values and assumptions	Make skillful use of reasoning and examples to broaden the context for analysis, support the thesis, and arrive at deeper insight into the issue Effectively convey reasons why the argument is worth considering Enrich and strengthen ideas and analysis by considering factors that complicate the writer's own perspective Anticipate objections by qualifying the argument	Group and sequence ideas logically, creating a progression that increases the effectiveness of the argument Use transitions between and within paragraphs to strengthen the relationships among ideas Make use of a controlling idea or purpose to unify and focus the argument	Make skillful and precise word choices that enhance the argument Make stylistic and register choices that are strategic and effective for the given writing purpose and topic Compose sentences with clear and consistently varied structures Produce writing that is free of all but a few minor errors in grammar, usage, and mechanics
9-10	Generate a precise thesis that establishes a perspective on a contemporary issue Engage productively with other perspectives on the issue Establish and employ a thoughtful context for analysis Address implications, complexities and tensions, and/or underlying values and assumptions	Make purposeful use of reasoning and examples to support the thesis and arrive at a deeper understanding of the issue Capably convey reasons why the argument is worth considering Enrich ideas and analysis by considering factors that complicate the writer's own perspective Anticipate objections by qualifying the argument	Group and sequence ideas logically to increase the effectiveness of the argument Use transitions between and within paragraphs to consistently clarify relationships among ideas Make use of a controlling idea or purpose to unify the argument	Make precise word choices that work in service of the argument Make stylistic and register choices that are effective for the given writing purpose and topic Compose sentences with clear and often varied structures Produce writing that has only minor errors in grammar, usage, and mechanics

ACT College and Career Readiness Standards—Writing—continued

	Ideas and Analysis (IOD)	Development and Support (SIN)	Organization	Language Use and Conventions
7-8	Generate a clear thesis that establishes a perspective on a contemporary issue Engage with other perspectives on the issue Establish and employ a relevant context for analysis Recognize implications, complexities and tensions, and/or underlying values and assumptions	Make use of clear reasoning and examples to arrive at an understanding of the issue and differing perspectives on it Adequately convey reasons why the argument is worth considering Extend ideas and analysis by considering factors that complicate the writer's own perspective Anticipate objections by qualifying the argument	Group and sequence ideas logically Use transitions between and within paragraphs to clarify relationships among ideas Make use of an emergent controlling idea or purpose to shape the argument	Make adequate word choices that convey the argument with clarity Make stylistic and register choices that are appropriate for the given writing purpose and topic Compose sentences with clear and occasionally varied structures Produce writing that has errors in grammar, usage, and mechanics but conveys meaning clearly
5-6	Generate a somewhat clear thesis that establishes a perspective on a contemporary issue Respond to other perspectives on the issue Establish a limited or tangential context for analysis Provide analysis that is simplistic or somewhat unclear	Make use of mostly relevant reasoning and examples to support the thesis and arrive at a general or simplistic understanding of the issue Offer a rationale that largely clarifies the argument Provide elaboration of ideas and analysis that is somewhat repetitive or imprecise	Group most ideas logically Use transitions between and within paragraphs to clarify some relationships among ideas Provide a basic organizational structure	Make word choices that are general and occasionally imprecise Make stylistic and register choices that are not always appropriate for the given writing purpose and topic Compose sentences that usually have clear structures but show little variety Produce writing that has distracting errors in grammar, usage, and mechanics but, in most instances, conveys meaning clearly

ACT College and Career Readiness Standards—Writing—continued

	Ideas and Analysis (IOD)	Development and Support (SIN)	Organization	Language Use and Conventions
3-4	Generate a thesis that is unclear or not entirely related to the given issue Respond weakly to other perspectives on the issue Provide analysis that is incomplete or largely irrelevant	Arrive at a weak understanding of the issue and differing perspectives on it through inadequate reasoning and examples Offer a rationale that fails to clarify the argument Provide elaboration of ideas and analysis that is illogical, disjointed, or circular	Group ideas with little consistency or clarity Use misleading and poorly formed transitions Provide a minimal organizational structure in which some ideas are grouped locally	Make word choices that are rudimentary and frequently imprecise Make stylistic and register choices that are inconsistent and are not always appropriate for the given writing purpose and topic Compose sentences that sometimes have clear structures Produce writing that has distracting errors in grammar, usage, and mechanics and only sometimes conveys meaning clearly
2	Scores below 3 do not permit useful generalizations about students' writing abilities.	Scores below 3 do not permit useful generalizations about students' writing abilities.	Scores below 3 do not permit useful generalizations about students' writing abilities.	Scores below 3 do not permit useful generalizations about students' writing abilities.

References

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

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

ACT. (2016b). *The ACT Aspire technical manual*. Iowa City, IA: Author.

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

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

Allen, J. (2013). *Updating the ACT College Readiness Benchmarks* (ACT Research Report No. 2013-6). Iowa City, IA: ACT, Inc.

Allen, J. & Sconing, J. (2005). *Using ACT Assessment scores to set benchmarks for college readiness* (ACT Research Report No. 2005-3). Iowa City, IA: ACT, Inc.

Camara, W. J., O’Connor, R., Mattern, K., & Hanson, M. A. (Eds.). (2015). *Beyond academics: A holistic framework for enhancing education and workplace success*. (ACT Research Report Series 2015-4). Iowa City, IA: ACT, Inc.

Mattern, K. D., Burrus, J., Camara, W. J., O’Connor, R., Gambrell, J., Hanson, M. A., Casillas, A., & Bobek, B. (2014). *Broadening the definition of College and Career Readiness: A Holistic Approach*. Iowa City, IA: ACT, Inc.

Mattern, K., Radunzel, J., & Westrick, P. (2015). *Development of STEM readiness benchmarks to assist educational and career decision making*. (ACT Research Report Series No. 2015-3). Iowa City, IA: ACT, Inc.

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

Radunzel, J., Westrick, P., Bassiri, D., & Li, D. (2017). *Development and validation of a preliminary ELA readiness benchmark based on the ACT ELA score.* (ACT Research Report No. 2017-9). Iowa City, IA: ACT, Inc.

Sawyer, R. (1989). *Validating the use of ACT Assessment score and high school grades for remedial course placement in college* (ACT Research Report No. 89-4). Iowa City, IA: American College Testing Program.

Chapter 9

Scaling and Equating

9.1 Overview

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 is described first, followed by the scaling and equating of the ACT writing test scores used for the ELA score calculation.

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

9.2.1 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 twelfth-grade college-bound examinees was used in scaling the ACT. A detailed discussion of the data used in the scaling of the ACT is given by Kolen and Hanson (1989).

The scaling process for the ACT consisted of three steps. First, weighted raw score distributions for both national and college-bound groups of examinees from the Academic Skills Study were computed. Second, the weighted raw score distributions were smoothed with a four-parameter beta compound binomial model (Lord, 1965; Kolen, 1991; Kolen & Hanson, 1989). Finally, the smoothed raw score distributions for twelfth-grade college-bound examinees were used to produce the score scales.

Smoothing the raw score distributions was done to produce distributions that are easier to work with and that are better estimates of 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 too much from the unsmoothed distributions. In addition, the first three central moments (mean, standard deviation, 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 next step in constructing the score scales was to produce initial scale scores with a specified mean and a specified conditional standard error of measurement (CSEM) that was approximately equal throughout the score scale for twelfth-grade college-bound examinees from the Academic Skills Study. Methods introduced by Kolen (1988) and described in detail by Kolen and Hanson (1989) were used for this process. These initial scale scores were rounded to integers ranging from 1 to 36 for the tests. Some adjustment of the rounded scale scores was performed to attempt to meet the specified mean and standard error of measurement (SEM) and to avoid gaps in the score scale (i.e., scale scores that were not used) or to avoid having too many raw scores convert to a single scale score.

In a special study 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.

9.2.2 Score Scale Characteristics

The scale score range is from 1 to 36 for the ACT multiple-choice tests as well as 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 among students at the beginning of twelfth grade, nationwide in 1988, who reported that they were planning to attend a two- or four-year college.

Although the score scale 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) have the same score range, scale scores on these two assessments are not directly comparable due to changes in the internal structure of the tests and the methodology used for scaling.

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. In addition, the scales for the ACT were constructed using a method described by Kolen (1988) 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 a reasonably good estimate of the measurement error at all score levels.

It should be noted that 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 reported score obtained over repeated administrations of the test under identical conditions. If one SEM were added to and subtracted from each of these reported scores, about 68% of the resulting intervals would contain the examinee's true score. This statement assumes a normal distribution for measurement error.

Another way to view 68% intervals is in terms of groups of examinees. Specifically, if one SEM were added to and subtracted from the reported score of each examinee in a group of examinees, the resulting intervals would contain the true score for approximately 68% of the examinees. To put it 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.

9.2.3 Equating

New forms of the ACT tests are developed each year. Even 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. As a result of this equating process, scale scores reported to examinees have the same meaning across all test forms and test dates.

A carefully selected sample of examinees from one of the national test dates each year is used as an equating sample in a randomly equivalent groups design. The examinees in this sample are administered a spiraled set of forms—the new forms and one anchor form that has already been equated to previous forms. 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 have the same percentile rank for a given group of examinees. The equipercentile equating results are smoothed using an analytic method described by Kolen (1984) to establish a smooth curve. The equivalents are then rounded to integers. The conversion tables that result from this process are used to transform raw scores on the new forms to scale scores.

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 a rounded arithmetic average of the scale scores from two or more tests. More information on these scores is provided in Chapter 7. The Composite, STEM, and ELA scores are also comparable across forms because the scores used to compute them have been equated.

9.3 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. Starting in 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 the calculation of 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 seven-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 ensured by the prompt selection procedures. Although prompts are selected to ensure that the 2–12 writing test scores are comparable no matter which prompt the student takes, that process does not ensure that the prompts are also strictly comparable for the sum of the four domain scores. Equating is used to adjust for slight differences in prompt difficulty for the sum of the domain scores that may still 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 postsmoothing under the randomly equivalent 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.

References

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.

Hanson, B. A. (1990). *An investigation of methods for improving estimation of test score distributions* (ACT Research Report No. 90-4). Iowa City, IA: American College Testing Program.

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

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

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

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: American College Testing Program.

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

Chapter 10

Reliability and Measurement Error

10.1 Overview

The potential for some degree of inconsistency or error is contained in 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 from a relatively small sample of items. In this chapter, a set of statistics are provided that quantify the reliability, measurement error, and classification consistency of the ACT test scores.

10.2 Reliability and Standard Error of Measurement

Reliability coefficients quantify the level of consistency of test scores. They typically 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. These coefficients are referred to as internal consistency reliability. Coefficient alpha (Cronbach, 1951) is one of the most widely used estimates of test reliability and was computed for the ACT tests. Coefficient alpha can be computed using the following formula

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

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

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

where SEM_t is the estimated scale score standard error of measurement and s_t^2 is the sample variance of the observed scale score for test *t*. The standard error of measurement (SEM) summarizes the amount of error or inconsistency in scores on a test. Scale score SEMs were estimated using a four-parameter beta compound binomial model as described in Kolen, Hanson, and Brennan (1992). If the distribution of measurement error is approximated by a normal distribution, true scale scores for about two-thirds of the examinees are within plus or minus one SEM from their reported scale score.

10.2.1 Reliability and SEM for the ACT Test Scores

Scale score reliability estimates and SEM for the four ACT multiple-choice tests (English, mathematics, reading, and science), Composite, STEM, and ELA scores are provided in Table 10.1. These values were calculated based on operational test data from five of the test forms administered in the 2015–2016 academic year. This is the same set of data used for analyses in Chapter 7. The reliability estimates are fairly high, with values over 0.9 for English, mathematics, Composite, STEM, and ELA scores, and values over 0.8 for reading and science. SEM values are fairly consistent across forms.

Table 10.1. Summary Statistics of Scale Score Reliability and SEM for the ACT Test Scores

		Reliability			SEM		
Test	# of Items	Median	Minimum	Maximum	Median	Minimum	Maximum
English	75	0.92	0.92	0.93	1.71	1.67	1.74
Mathematics	60	0.91	0.90	0.92	1.55	1.45	1.63
Reading	40	0.87	0.85	0.88	2.16	2.07	2.27
Science	40	0.85	0.81	0.85	2.01	1.90	2.16
Composite	215	0.97	0.96	0.97	0.93	0.93	0.96
STEM	100	0.93	0.93	0.94	1.29	1.24	1.30
ELA	116	0.91	0.88	0.91	1.59	1.58	1.60

10.2.2 Reliability and SEM for ACT Reporting Category Scores

Raw score reliability estimates, computed using coefficient alpha, and SEM were also calculated for the ACT reporting categories. These values, provided in Table 10.2, were calculated using operational test data from 11 forms administered in the 2015–2016 academic year. For some of the reporting categories, particularly those with very few items, the reliability is low. However, reporting category scores are not intended for high-stakes decisions. They are intended to guide instruction and help identify students' strengths and weaknesses.

Table 10.2. Summary Statistics of Raw Score Reliability and SEM for the ACT Reporting Categories

Test/Reporting Categories	Median # of Items	Reliability			SEM		
		Median	Minimum	Maximum	Median	Minimum	Maximum
English							
Production of Writing	23	0.81	0.77	0.82	2.07	1.99	2.12
Knowledge of Language	12	0.67	0.63	0.72	1.46	1.36	1.52
Conventions of Standard English	40	0.86	0.84	0.88	2.66	2.58	2.74
Mathematics							
Preparing for Higher Math	35	0.84	0.80	0.86	2.51	2.47	2.56
Number & Quantity	5	0.33	0.26	0.54	0.94	0.89	1.00
Algebra	8	0.53	0.49	0.65	1.19	1.13	1.23
Functions	8	0.59	0.48	0.65	1.20	1.15	1.24
Geometry	8	0.55	0.48	0.60	1.22	1.13	1.26
Statistics & Probability	6	0.42	0.34	0.51	1.06	1.01	1.09
Integrating Essential Skills	25	0.81	0.77	0.84	2.11	2.07	2.17
Modeling	24	0.80	0.71	0.84	2.09	1.73	2.32
Reading							
Key Ideas & Details	23	0.78	0.74	0.80	2.10	2.02	2.19
Craft & Structure	11	0.60	0.54	0.65	1.45	1.32	1.54
Integration of Knowledge & Ideas	6	0.44	0.34	0.55	1.10	0.82	1.14
Science							
Interpretation of Data	18	0.72	0.63	0.75	1.77	1.61	1.95
Scientific Investigation	10	0.58	0.47	0.69	1.39	1.14	1.71
Evaluation of Models, Inferences & Experimental Results	12	0.64	0.45	0.74	1.50	1.28	1.69

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

Whereas the SEM provides an average measure of score variability (or unreliability) 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 any particular ACT test score is approximately the same for low-scoring examinees as it is for high-scoring examinees.

For the ACT, the CSEMs were computed using methods described by Kolen, Hanson, and Brennan (1992). Figure 10.1 presents the CSEMs for the four multiple-choice tests across five of the forms administered in the 2015–2016 academic year. The CSEM is not graphed 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 are expected to have a true scale score lower than the minimum plotted score for each administration.

For most of the true scale score range, the scale score CSEM is reasonably constant. Some deviations occur at higher true scale scores. Some of these deviations are due to gaps in the raw-to-scale-score conversion at the high end of the scale for certain forms (for some forms certain scale scores cannot be obtained at the high end of the scale). For all tests, the CSEM is smaller at very high scores. The CSEM must be zero for the maximum true scale score and be near zero for true scale scores near the maximum. For this reason, the method used to produce the score scales cannot guarantee a completely constant CSEM for all true scale scores. However, the proportion of examinees with true scores at the extreme high end of the scale is very low. For the vast majority of examinees, the constant CSEM property is reasonably well met.

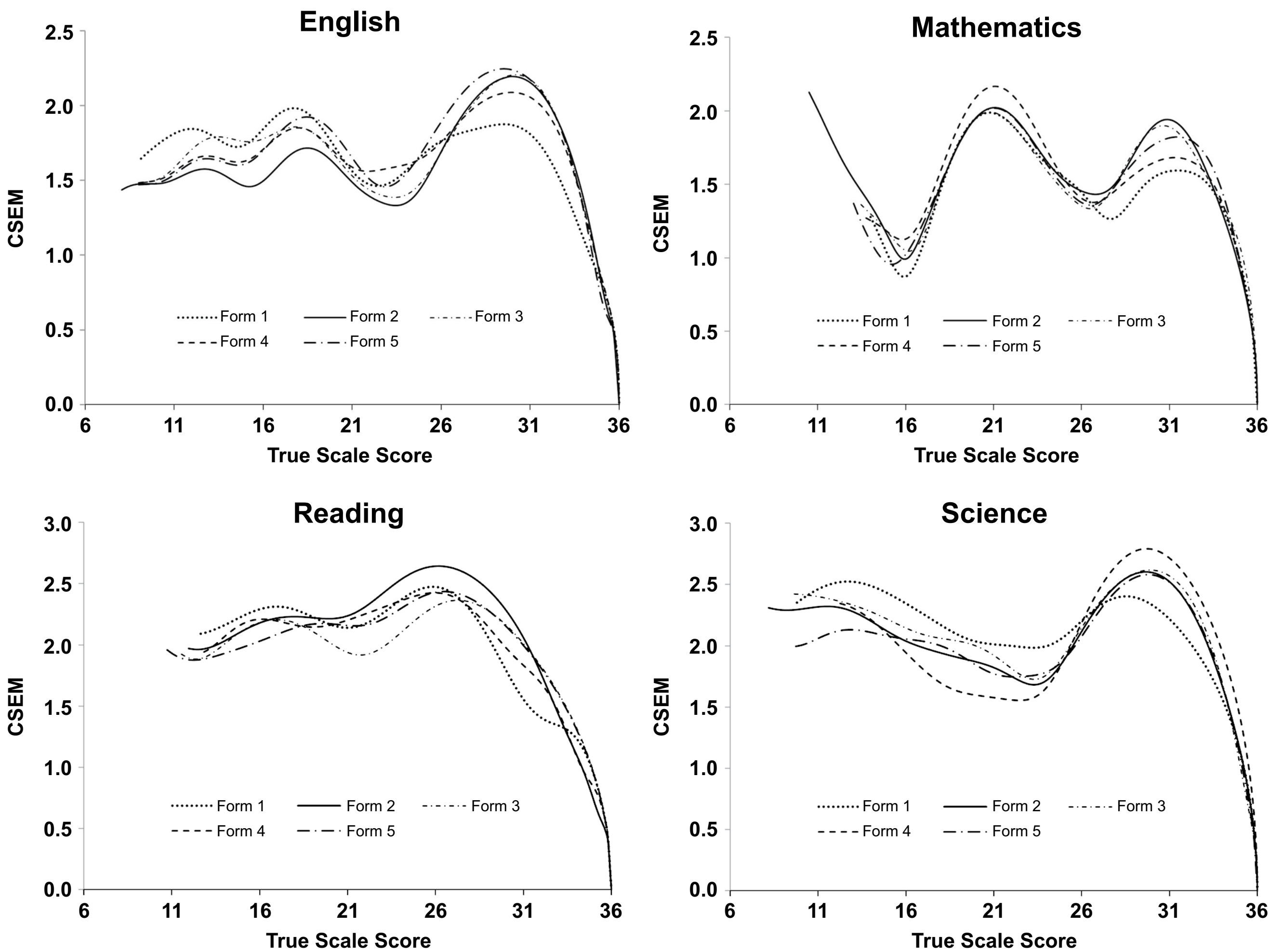


Figure 10.1 CSEM for multiple-choice test scores.

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

Estimates of reliability and CSEMs for the writing test were computed 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 different parts. Each part involved a different pair of schools. Within each pair of schools, two writing prompts were used. The responses to both writing prompts were rated by three raters on the four different 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 different replications of the same study design. The estimated variance components for the interactions between both prompt and rater as well as person and rater were quite small across all three school pairs. The estimated variance components for the interaction between person and prompt were relatively large for all three pairs, however. This is consistent with results typically found in the literature. 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.

To estimate the reliability and SEM for writing scores on the 1–36 scale, data from the 2014 writing field test study were used. 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 quadratic polynomial. The square root of these fitted values is represented by the solid line in Figure 10.2. 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 0.68 and the scale score SEM was 3.89. This SEM value was used in the calculation of the ELA reliability and SEM.

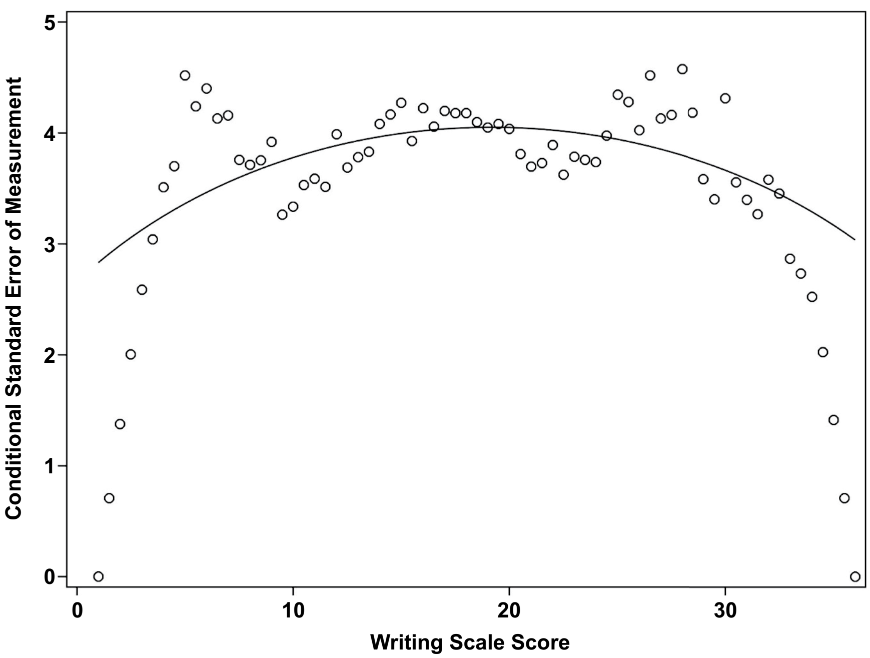


Figure 10.2 Average and fitted CSEMs for ACT writing test scale scores.

Operational agreement indices were also calculated based on five of the forms administered in the 2015–2016 academic year. As shown in Table 10.3, these agreement indices included the perfect agreement rate, the perfect plus adjacent agreement rate, and the quadratic weighted kappa coefficient. The perfect agreement rate, or percent of students who received the same domain score (from 1 to 6) from both raters, ranged from approximately 0.50 to 0.57 across domains and forms. The perfect plus adjacent agreement rates, or the percent 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 both raters, was very high, ranging from approximately 0.92 to 0.97 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 reflect the relative difference between categories. The kappa coefficient is a positive number if the observed agreement is larger than the chance agreement, with larger numbers representing more agreement between two raters. Fleiss, Levin, and Paik (2003) indicated that for most purposes, kappa values larger than 0.75 may represent excellent agreement beyond chance, values below 0.40 may represent poor agreement beyond chance, and values in between may represent fair to good agreement beyond chance. The quadratic weighted kappa coefficients for the ACT writing domain scores ranged from 0.58 to 0.66, indicating good rater agreement.

Table 10.3. Agreement Rates for the ACT Writing Domain Scores

Domain	Agreement Index	Median	Minimum	Maximum
Ideas & Analysis	Perfect Agreement	0.53	0.50	0.55
	Perfect + Adjacent Agreement	0.94	0.92	0.97
	Quadratic Weighted Kappa	0.62	0.61	0.65
Development & Support	Perfect Agreement	0.52	0.50	0.56
	Perfect + Adjacent Agreement	0.94	0.92	0.97
	Quadratic Weighted Kappa	0.62	0.61	0.66
Organization	Perfect Agreement	0.53	0.51	0.57
	Perfect + Adjacent Agreement	0.95	0.93	0.97
	Quadratic Weighted Kappa	0.61	0.60	0.65
Language Use & Conventions	Perfect Agreement	0.54	0.51	0.56
	Perfect + Adjacent Agreement	0.95	0.93	0.97
	Quadratic Weighted Kappa	0.59	0.58	0.61

10.2.5 CSEM for Composite Scores

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

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

where $s_i(\tau_i)$ is the CSEM for test i at true scale score τ_i and $i = e, m, r$, and s for English, mathematics, reading, and science, respectively. The functions $s_i(\tau_i)$ are plotted in Figure 10.1. The CSEM for the Composite score is plotted as a function of the average of the true scale scores variances for the four tests. 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 plots of the CSEMs for the Composite score, the observed proportion-correct scores (the number of items correct divided by the total number of items) for examinees on the four tests were treated as true proportion-correct scores at which the CSEMs were calculated. For each test the CSEM was computed for each examinee using the observed proportion-correct score as the true proportion-correct score in the formula for the CSEM (Equation 8 in Kolen, Hanson, & Brennan, 1992). In addition, for each test the true scale score corresponding to the observed proportion-correct score (treated as a true proportion-correct score) was computed (Equation 7 in Kolen, Hanson, & Brennan, 1992). The resulting CSEMs for the four tests were substituted in the equation given above to compute the CSEM for the Composite score. The CSEM for the Composite score was plotted in Figure 10.3. This procedure was repeated for each of the examinees from five of the test forms administered in the 2015–2016 academic year. Values for examinees who received proportion-correct scores of 0 or 1 on any of the four tests are not plotted in Figure 10.3. While observed proportion-correct scores of 0 and 1 are possible, true proportion-correct scores of 0 and 1 are unrealistic.

The CSEMs presented in Figure 10.3 vary not only across Composite scale scores but also within each Composite scale score. Different CSEMs are possible for each particular value of the Composite scale score because more than one combination of the four test scores can produce the same average scale score. The general trend in the plots is that the CSEMs are fairly constant in the middle of the scale and lower for moderately high scores. This trend is similar to the trend in Figure 10.1 for the CSEM for the four tests. The CSEM of the Composite score is, for practical purposes, reasonably constant across the score scale.

A limitation of the approach used in producing the CSEM estimates of the Composite score in Figure 10.3 is that they correspond to the unrounded average of the four test scores rather than to the rounded average of the four test scores, which is the Composite score reported to examinees.

However, it is not a problem that the observed scores of the examinees are used in producing the plots because it is the standard errors *conditional* on average true scale score that are being plotted, and the observed scores for the examinees are only used to determine the specific average true scale scores at which to plot the CSEMs. One effect of using observed scores as the true score values at which to plot the CSEM is that many points at the extremes of the scale in Figure 10.3 may not represent realistically obtainable average true scale scores since the probability of observing examinees with these values of average *true* scale scores is extremely small.

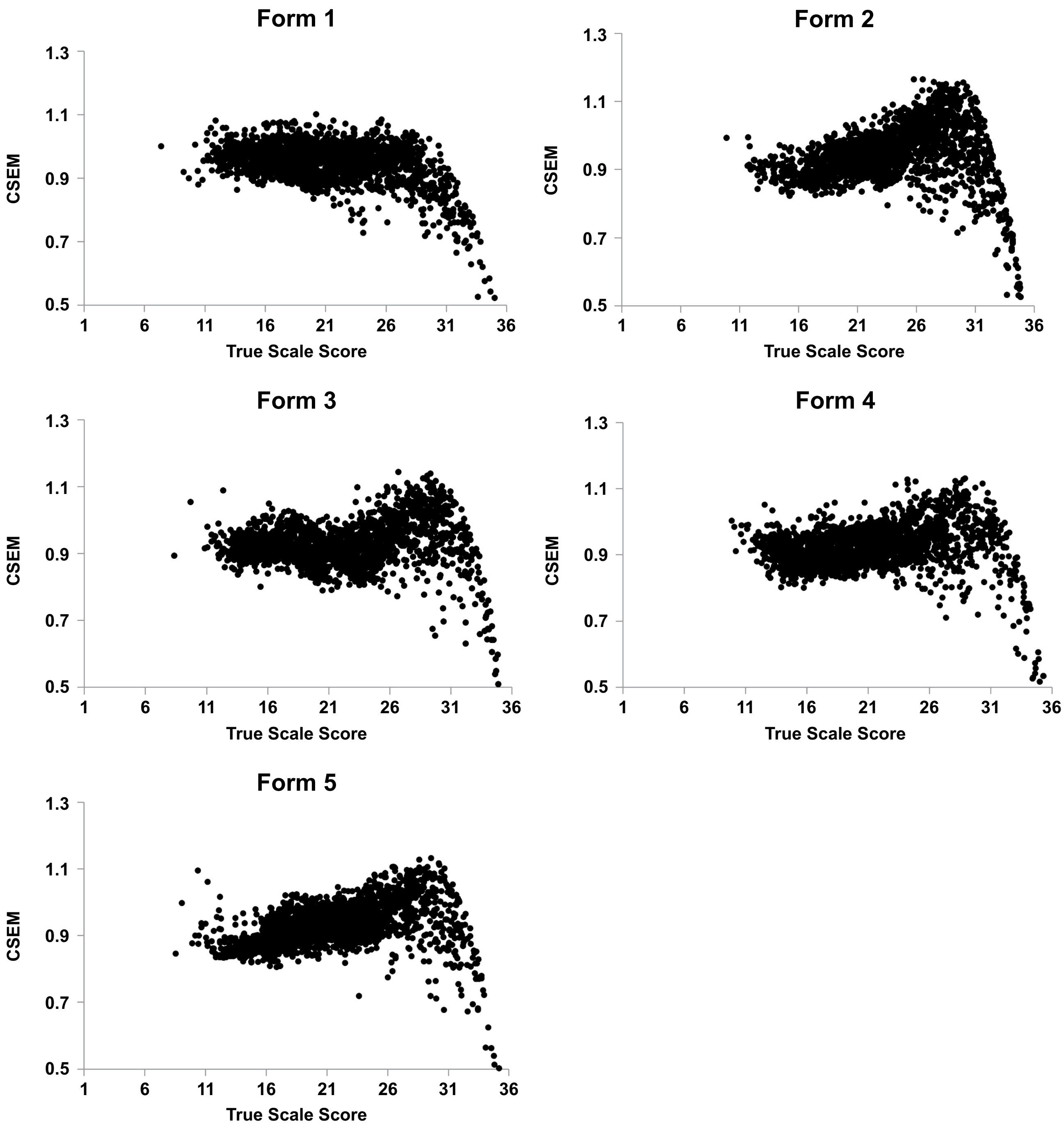


Figure 10.3 CSEM for Composite scores.

10.2.6 CSEM for STEM and ELA Scores

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

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

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

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

where $s_i(\tau_i)$ is the CSEM for test i at true scale score τ_i and $i = e, r,$ and w for English, reading, and writing, respectively. The same set of data used to produce the CSEM values for the Composite score was used to obtain the CSEM values for the STEM scores plotted in Figure 10.4 and the CSEM values for the ELA scores in Figure 10.5.

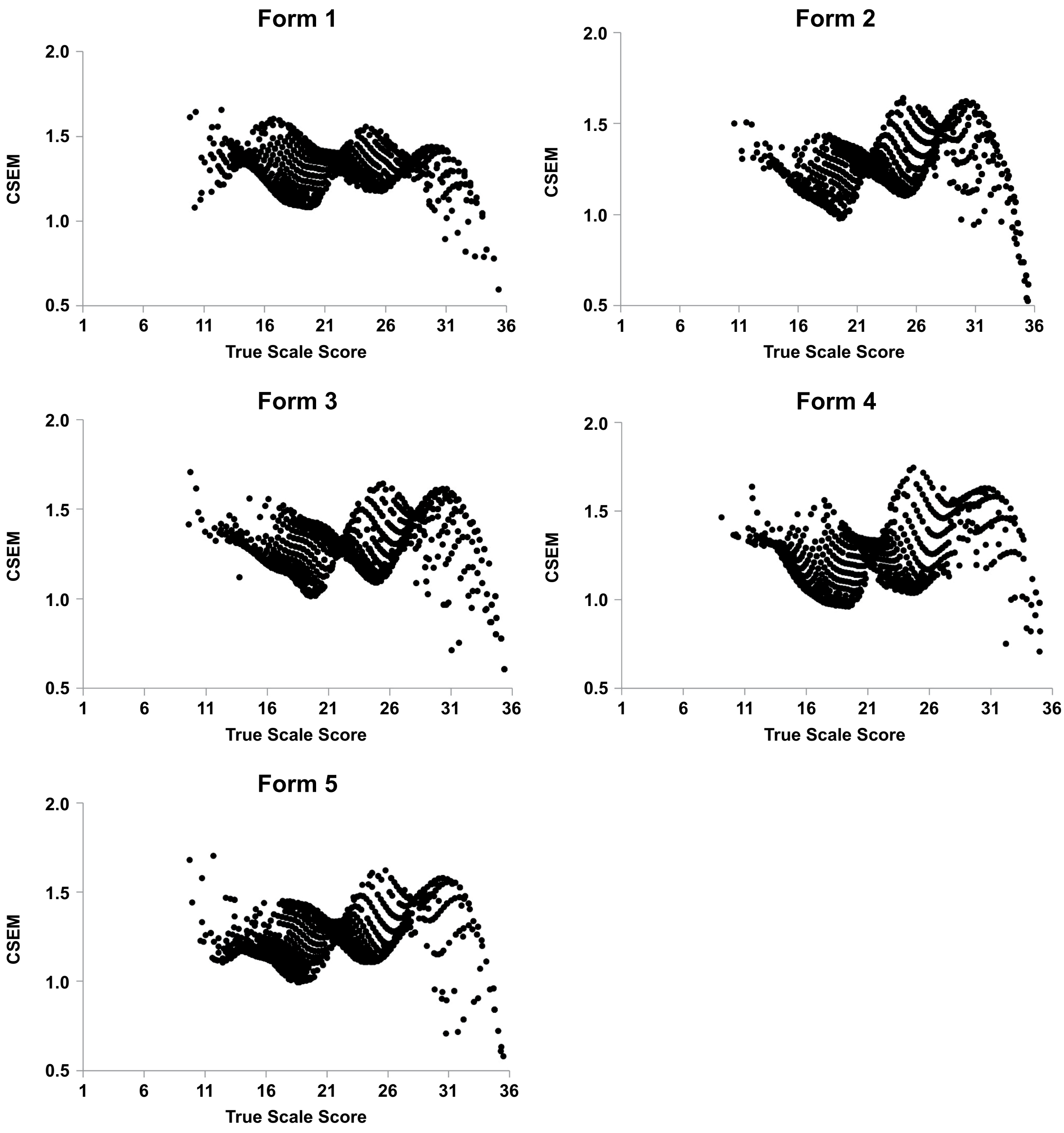


Figure 10.4 CSEM for STEM scores.

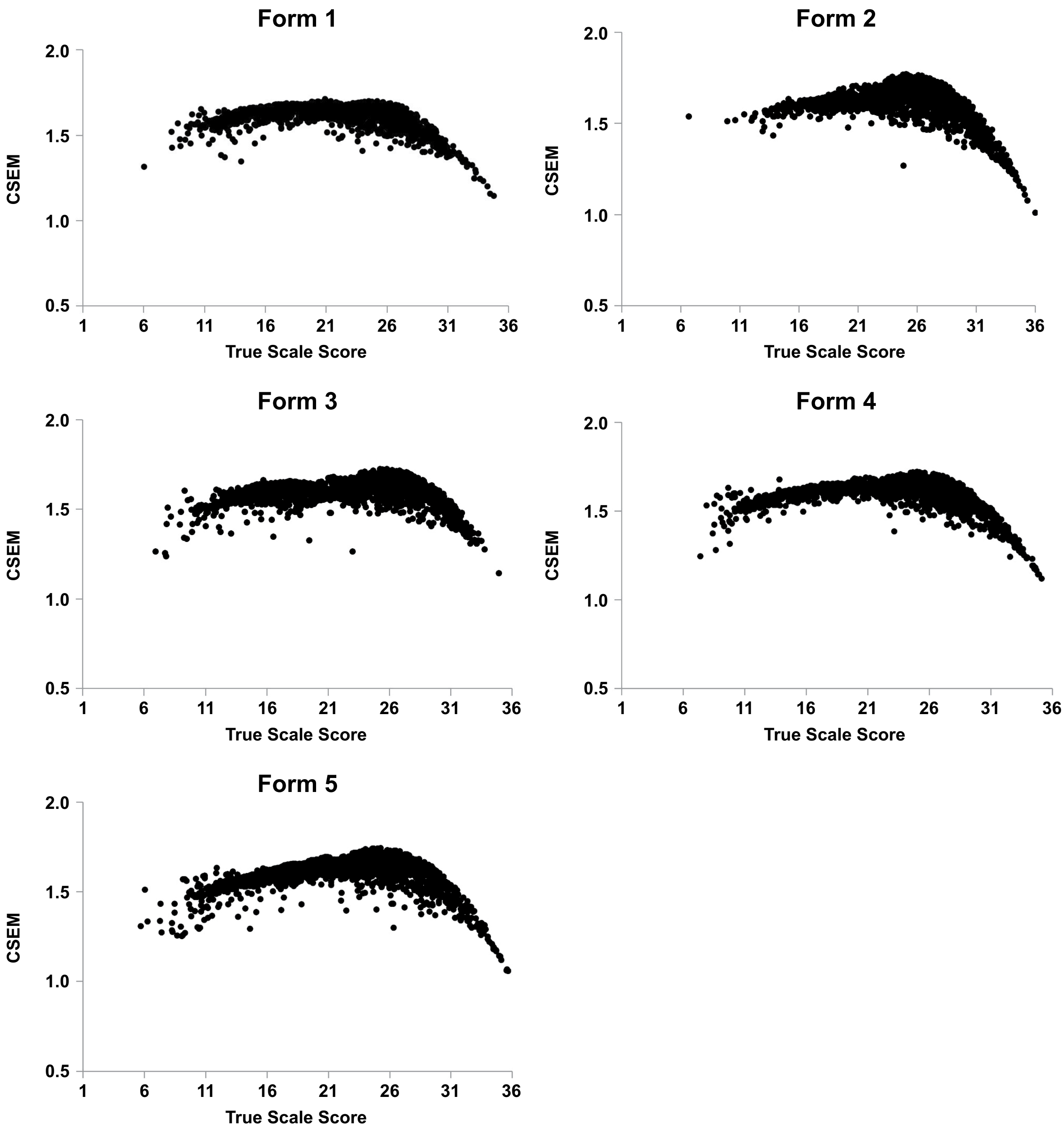


Figure 10.5 CSEM for ELA scores.

10.3 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.

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

Classification consistency values were computed using data from five forms administered in the 2015–2016 academic year for the four ACT multiple-choice tests and the STEM and ELA scores. Classification was based on the ACT College Readiness Benchmarks (see Chapter 8 for detail about the Benchmarks). The classification consistency results are provided in Table 10.4. Values are all fairly high, ranging from a low of 0.83 in science to a high of 0.94 for STEM.

Table 10.4. Classification Consistency for the ACT Readiness Benchmarks

		Classification Consistency		
Test	Number of Items	Median	Minimum	Maximum
English	75	0.89	0.88	0.90
Mathematics	60	0.89	0.87	0.91
Reading	40	0.86	0.86	0.86
Science	40	0.85	0.83	0.87
STEM	100	0.93	0.90	0.94
ELA	116	0.88	0.87	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 10.5, are based on data from 11 forms administered during the 2015–2016 school year.

10.3.2 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 five of the forms administered in the 2015–2016 academic year, the classification consistency ranged from 0.64 to 0.69, which was moderately high considering the number of items that contribute to UCT scores and the number of performance levels. The number of UCT items ranged from 16 to 21 across these five forms, and the percentages of students classified as Below Proficient, Proficient, and Above Proficient were 43%, 33%, and 24%, respectively.

10.3.3 Classification Consistency for Progress Toward ACT NCRC Indicator

The second indicator, Progress Toward the ACT National Career Readiness Certificate (ACT NCRC), had classification consistency values ranging from 0.78 to 0.80 across five of the forms administered in the 2015–2016 academic year. These values are quite high given that there are four performance levels for the ACT NCRC, as shown in Table 10.6. Note that the classification consistency index is an indication of the stability of the Progress Toward ACT NCRC indicator if different ACT test forms were taken and is not an indication of the accuracy of the classification compared with students' actual NCRC attainment.

Table 10.5. Classification Consistency for the ACT Readiness Ranges

Test/Reporting Categories	Median # of Items	Classification Consistency		
		Median	Minimum	Maximum
English				
Production of Writing	23	0.82	0.79	0.83
Knowledge of Language	12	0.76	0.75	0.82
Conventions of Standard English	40	0.85	0.83	0.86
Mathematics				
Preparing for Higher Math	35	0.83	0.81	0.85
Number & Quantity	5	0.61	0.58	0.72
Algebra	8	0.68	0.66	0.75
Functions	8	0.72	0.67	0.75
Geometry	8	0.69	0.66	0.73
Statistics & Probability	6	0.66	0.61	0.70
Integrating Essential Skills	25	0.81	0.79	0.84
Modeling	24	0.81	0.76	0.83
Reading				
Key Ideas & Details	23	0.80	0.78	0.82
Craft & Structure	11	0.73	0.70	0.76
Integration of Knowledge & Ideas	6	0.67	0.60	0.71
Science				
Interpretation of Data	18	0.77	0.73	0.79
Scientific Investigation	10	0.71	0.65	0.76
Evaluation of Models, Inferences & Experimental Results	12	0.74	0.68	0.80

Table 10.6. Composite Score Ranges for the ACT NCRC Levels

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

References

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.

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

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.

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

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.

Chapter 11

Validity Evidence for the ACT Tests

11.1 Overview

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 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.

11.2 Measuring 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 nine subsections and provides validity evidence for using ACT test scores to measures students' educational achievement. The first subsection provides a content validity argument for ACT scores. The next five subsections focus on relating high school course work, grades, end-of-course exams, and noncognitive factors to ACT scores and ACT Benchmark attainment. The seventh subsection focuses on understanding subgroup differences on the ACT. The eighth subsection focuses on the relationships between test preparation activities and ACT performance. The ninth subsection focuses on the use of ACT scores for measuring educational achievement for gifted and talented programs.

11.2.1 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 assessment 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.[1] 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 than are courses taken and grades earned.

11.2.2 Statistical Relationships between ACT Scores and High School Course Work and Grades

The ACT tests are oriented toward the general content areas of high school and college curricula. 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 entails the completion of the Course/Grade Information Section (CGIS), which collects information about 30 high school courses in English, mathematics, social

[1] 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).

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 course work 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 11.1 displays the ACT scale score means and standard deviations in English, mathematics, reading, and science tests for three groups of students by years of English, mathematics, social studies, and science course work expected to complete in high school (based on courses identified as taken or plan to take on the CGIS; 7% of the students were missing this information). For the ACT English test, the largest score differences are, not unexpectedly, between those who expected to take at least 3½ years of English and those who expected to take 2 years or less. This pattern is also apparent for the ACT mathematics, reading, and science tests. These findings are similar to those found in an earlier study based on a nationally representative sample (Harris & Kolen, 1989).[2]

Table 11.1 Means and Standard Deviations for ACT Scores: 2016 ACT-Tested High School Graduates by Years of Subject-Relevant Course Work

Years of course work	English			Mathematics			Reading			Science		
	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD
≤ 2	24,520	14.2	5.4	71,009	15.8	3.0	96,526	17.8	6.1	255,155	18.0	4.9
2½–3	73,927	15.5	5.6	287,611	17.0	3.3	373,246	20.6	6.6	835,004	20.8	5.3
> 3	1,844,583	20.7	6.7	1,577,398	21.7	5.4	1,467,317	22.0	6.4	844,557	22.1	5.7

Moreover, as shown in Table 11.2, 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.

[2] The Harris and Kolen (1989) study examined just the relationships between years of English and mathematics course work and ACT English and mathematics scores.

Table 11.2 Average ACT Scores by Academic Preparation, 2012–2016

Academic preparation	Reference year	N	ACT score				
			English	Mathematics	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

*Core curriculum is defined here 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 11.1 and 11.2 support the notion that the ACT is a curriculum-based test. Additionally, an analysis by McNeish, Radunzel, and Sanchez (2015) showed that, in general, course work and high school grades were strongly associated with performance on the ACT, after statistically controlling for other factors. However, it is also conceivable that some other factors, to include noncognitive factors, account for the observed association between high school course work and ACT scores. In the McNeish et al. study, the researchers investigated the relationships between noncognitive characteristics, high school course work and grades, school characteristics, and test scores of ACT-tested students. The reminder of this section describes this study in detail.

Data. A random sample of 56,000 high school seniors who registered for the ACT in either October or December of 2012 was invited to complete an online questionnaire on the Monday after the date of the ACT test administration. The questionnaire asked students about their high school experience, study and work habits, parental involvement, educational and occupational plans and goals, and college courses taken and college credits earned in high school. The final sample consisted of 6,440 high school seniors from 4,541 high schools who took the ACT in the fall of 2012 and completed the online questionnaire. Twelve percent of the initial sample responded and met the study inclusion criteria.

Method. A blockwise regression model with cluster-robust standard errors was used to model five ACT test scores (English, mathematics, reading, science, and Composite) using high school course work and grades, school characteristics, and noncognitive variables. Related predictor variables were grouped in blocks, and the blocks were added one at a time to examine incremental improvements to the variance explained by the regression model (see Table 11.3 for the various block groupings denoted in bold font; results for gender and race/ethnicity are shown in Table 11.16). A stepwise selection procedure was employed within each block. To be retained in the models, variables within the blocks were required to have statistically significant regression coefficients ($p < .01$). The blocks were entered into the regression model in the following order: high school course grades, course work taken, advanced course work taken, school characteristics, noncognitive characteristics, socioeconomic status (SES)-related demographics, gender, and race/ethnicity. Upon entry, the contribution of each variable block was evaluated relative to the blocks preceding it; this procedure continued until all blocks were evaluated. Once a predictor was included based on the statistical significance of its regression coefficient, it was retained in the model regardless of whether the *p* value changed after subsequent blocks were added. Weighted analyses were utilized to ensure that the sample resembled the population in terms of student demographics and achievement levels. For a more comprehensive description of the methods and online questionnaire, see ACT Research Report No. 2015-6 (McNeish et al., 2015).

Results. Multiple regression statistics for modeling ACT scores are reported in Table 11.3. Regression coefficients, total R^2, and the root mean square error (RMSE) are reported by model for each ACT score. High school grade point average (HSGPA) accounted for a larger percentage of the variance in ACT scores than any other predictor in the model (20% to 31%; Figure 11.1). The mathematics and science course sequence taken accounted for an additional statistically significant proportion of the variance in ACT scores (from 4% to 13%). This is not to say that other courses taken, including English and social studies, were unrelated to ACT performance. In general, the other courses taken were collinear with mathematics and science courses, or they had little variance (i.e., most students took or did not take these courses). Taking advanced high school course work, such as accelerated, advanced, honors, or courses for college credit, accounted for an additional 3% to 5% of the variance in ACT scores. HSGPA and course work taken, in combination, explained between 28% and 46% of the variance in ACT scores. After all blocks were entered, the models for the ACT mathematics score and Composite score had the greatest prediction accuracy based on total R^2 (.60 and .61, respectively). That is, 60% to 61% of the variance in ACT mathematics and Composite scores could be explained by the predictors in the model. The percentage of variance explained was lower for ACT English scores (56%), ACT science scores (49%) and ACT reading scores (44%).

The individual unstandardized regression coefficients reported in Table 11.3 can be interpreted as the expected change (increase or decrease) in ACT scores associated with the predictor, holding the other variables in the model constant. For example, as shown in Table 11.3, taking higher-level mathematics courses beyond Algebra 2 was associated with an average ACT mathematics test score increase of 0.7 to 3.0 scale score point, compared to taking a mathematics sequence that included Algebra 1, Geometry, and Algebra 2. For the science course sequence, taking Biology, Chemistry, and Physics was associated with average ACT score increases of 0.5 to 0.8 scale score point on the ACT mathematics and science tests and the Composite, compared to taking Biology only. Controlling for the other variables in the models, students taking advanced course work in English were expected to score 1.0 to 1.1 points higher on the ACT reading and English tests. In contrast, taking advanced course work in English was not related to performance on the ACT mathematics and science tests.

Summary. In this study, between 44% and 61% of the variance in ACT scores was explained by HSGPA, course work taken, school characteristics, noncognitive characteristics, and demographic characteristics. High school academic factors, such as HSGPA and course work, accounted for the most variance explained in all five ACT scores (R^2 = 0.28 to 0.46). The first three blocks comprised 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 course work 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 course work, grades, and ACT score relationships.

Table 11.3 Weighted Regression Statistics for Modeling ACT Scores

	ACT score				
Predictor	English	Mathematics	Reading	Science	Composite
Intercept	17.73	20.14	20.59	20.45	19.80
HSGPA in 4 core areas[a]	2.74	2.05	2.16	1.83	2.18
High school course information					
Mathematics course sequence[b]					
Less than Alg 1, Geom, Alg 2	–0.41*	–0.39**	–0.25*	–0.69**	–0.38**
Alg 1, Geom, Alg 2 (referent)					
Alg 1, Geom, Alg 2, Other Adv. Math	0.58	0.71	0.57	0.56	0.59
Alg 1, Geom, Alg 2, Trig	0.64	0.82	0.40	0.41**	0.54
Alg 1, Geom, Alg 2, Other Adv. Math, Trig	1.57	1.63	1.10	1.21	1.33
Alg 1, Geom, Alg 2, Trig, Calc	2.04	2.62	1.68	2.01	2.04
Alg 1, Geom, Alg 2, Other Adv., Trig, Calc	2.37	3.02	1.86	2.21	2.32
Other math sequence of 3 or more years	0.94*	1.59	0.50*	1.18	0.99
Other math sequence of less than 3 years	0.58*	0.77*	0.38*	0.28**	0.56**
Science course sequence					
Less than Biology[c]	0.58*	0.78*	—	0.40*	0.48*
Biology (referent)					
Biology and Chemistry	0.39**	0.34**	—	0.18**	0.27**
Biology, Chemistry, and Physics	0.39**	0.82	—	0.60	0.53
Other science sequence	–0.08*	0.55**	—	0.07*	0.12*
Years of foreign language	0.10**	—	—	—	—
Advanced high school course work					
Advanced English (taken/not taken)[d]	1.13	–0.15**	0.99	—	0.54
Advanced mathematics (taken/not taken)[d]	—	1.30	—	0.68	0.66

Table 11.3 Weighted Regression Statistics for Modeling ACT Scores—continued

	ACT score				
Predictor	English	Mathematics	Reading	Science	Composite
Advanced natural science (taken/not taken)[d]	0.67	0.63	0.42	0.64	0.49
Advanced social studies (taken/not taken)[d]	1.10	0.30	1.12	0.40	0.69
College credits earned in high school					
0 (referent)					
1 to 6	–0.12*	0.26††	–0.09*	–0.03*	–0.04*
7 or more	0.26*	0.60	0.42††	0.44	0.39
High School characteristics					
Median zip code income					
Low [< $35,421] (referent)					
Middle [$35,421–$47,852]	0.41	0.46	0.47	0.53	0.48**
High [> $47,852]	0.60	0.70	0.53	0.72	0.67
% college enrollment	—	0.01	—	0.01	—
% free/reduced lunch					
Low [< 25%] (referent)					
Middle [25%–50%]	–0.27**	–0.37	–0.28*	–0.15*	–0.27*
High [> 50%]	–0.59	–0.59	–0.44**	–0.33**	–0.51
% intending graduate degree	0.03	0.02	0.03	0.01*	0.03
Quadratic term	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
% minority					
Low [< 9%] (referent)					
Middle [9%–36%]	–0.15**	–0.23**	–0.14**	–0.09**	–0.16**
High [> 36%]	–0.87	–0.78	–0.93	–0.78	–0.87
Non-public school indicator	0.70**	–0.76	0.15*	–0.69	–0.13*
Noncognitive characteristics					
College prep course curriculum (taken/not taken)	0.41	—	0.47	0.28††	0.34
Educational aspirations					
Below bachelor's (referent)					
Bachelor's degree	0.50*	0.24*	0.29*	0.28*	0.34*
Beyond bachelor's degree	1.34	0.81	1.21	0.92	1.08

Table 11.3 Weighted Regression Statistics for Modeling ACT Scores—continued

	ACT score				
Predictor	English	Mathematics	Reading	Science	Composite
Need help with educational/occupational plans	0.38	—	—	—	—
Need help with writing skills (yes/no)	–0.26**	—	—	—	—
Need help with study skills (yes/no)	–0.34††	—	—	—	—
Need help with reading (yes/no)	–1.69	—	–2.39	–0.94	–1.33
Need help with math skills (yes/no)	—	–1.49	—	–0.69	–0.52
Parents check assignments	–0.41	–0.24	–0.35	–0.23	–0.31
Perception of education (PCA component)	—	0.16	—	0.19	0.13
Student challenged by school	–0.41	–0.27	–0.49	–0.36	–0.39
Tested in junior year	1.35	0.58	0.64	0.74	0.77
SES-Related Demographics					
English spoken at home	0.99	—	0.91	0.68	0.70
Family income					
< $36,000 (referent)					
$36,000 to $80,000	0.37††	0.16**	—	0.22**	0.24**
> $80,000	0.61	0.46	—	0.26**	0.39
Highest parental educational level					
No college (referent)					
Some college	0.56	0.15*	0.54	0.21*	0.36
Bachelor's degree	0.91	0.35**	0.89	0.34**	0.61
Graduate degree	1.14	0.35**	1.11	0.44**	0.73
Total R^2	.56	.60	.44	.49	.61
Mean square error	4.22	3.21	4.47	3.54	3.13

Note. Regression coefficients for all achievement, school characteristics, and noncognitive variables were statistically significant ($p < .01$) unless denoted otherwise. Regression coefficients for gender and race/ethnicity are shown in Table 11.16.

† indicates a *p*-value between 0.010 and 0.015 upon entry to final model.

†† indicates a *p*-value between 0.010 and 0.015 in the final model.

* indicates that the indicator was not statistically significant upon entry but was retained as part of a predictor.

** indicates that the predictor was statistically significant upon entry but was no longer significant in the final model.

[a]Average of course grades in 23 core courses in English, mathematics, natural sciences, and social studies. This variable was grand-mean centered at 3.31.

[b]Alg = Algebra; Geom = Geometry; Oth. Adv. Math. = other advanced math course beyond Algebra 2; Trig = Trigonometry; Calc = Calculus.

[c]Sample size for the less than Biology course sequence was relatively small (< 100 students).

[d]Advanced course work includes any accelerated, advanced, honors, and dual-enrollment courses taken in the subject area by the student while in high school.

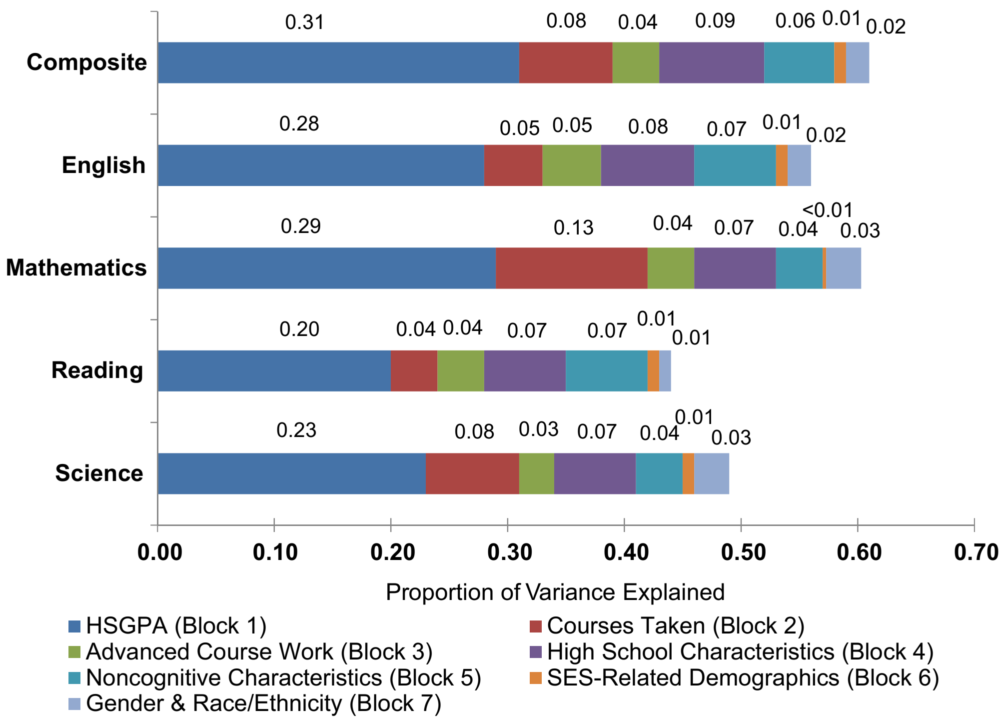


Figure 11.1 Proportion of variance in ACT scores associated with HSGPA, high school course work taken, school characteristics, noncognitive characteristics and demographic characteristics (McNeish et al., 2015).

11.2.3 Construct Contamination in HSGPA

ACT scores are statistically associated with high school grades (Table 11.4; see also the previous section). Students who have higher HSGPAs tend to achieve higher ACT scores. 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 the differing standards and purposes teachers associate with grades (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 validity of high school grades.

Table 11.4 Average ACT Score by HSGPA Ranges, 2015–2016

		ACT score									
		English		Mathematics		Reading		Science		Composite	
Group	N	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
All students	2,090,342	20.1	6.8	20.6	5.4	21.3	6.5	20.8	5.6	20.8	5.6
HSGPA:											
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
1.99 and below	75,255	13.6	4.3	15.7	2.5	15.4	4.2	15.5	3.9	15.2	3.1

Reliability of Self-Reported Course Work and Grades. The accuracy of the high school course and grade information students provide in the ACT registration folder within 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 course work and grades differed little by gender, race/ethnicity, and income. The results indicated that self-reported course work 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 achievement. A study by Woodruff and Ziomek (2004a) investigated inflation in HSGPA; this study was a follow-up to an earlier study by Ziomek and Svec (1995). The latter study examined ACT Composite scores and HSGPAs from 1990 to 1994 and found evidence for modest grade inflation. The results from the former study (1991–2004) 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. The reminder of this section describes this study in detail.

Data and method. The data for the Zhang and Sanchez (2013) study included public high school graduates from 2004 to 2011 who took the ACT test in the eleventh or twelfth grade of high school as a part of national testing or a statewide adoption program. High schools were included in the analysis if they had at least 100 ACT-tested students across the eight years examined. If a student took the ACT test more than once, the most recent test record was used. High school grades in up to 30 courses were self-reported by students when they registered to take the ACT test. Overall HSGPA was calculated based on grades in 23 of the 30 courses from the CGIS; grades in foreign language and art courses were not included. Student-level data were aggregated at the school level to explore school-level grade inflation. Conditional average HSGPAs were calculated by ACT Composite score for each high school and each year. For these analyses, the school was the unit of analysis.

Results. Table 11.5 shows the number of high schools by year from 2004 to 2011, as well as the average HSGPA and demographic variables across high schools. The state-tested population percentage also increased by approximately 10 percentage points during that period. This is partially a result of additional states incorporating the ACT test into their statewide high school assessment programs. The average HSGPA and average ACT Composite (ACT-C) score for schools were similar across years, which suggested that grade inflation may not be observed in the period from 2004 to 2011. The average free/reduced-price lunch eligible percentage and the racial/ethnic minority percentage were also consistent across the eight years examined.

The curves in Figure 11.2 show simple averages across high schools of the conditional mean HSGPAs, given ACT-C score. There is a separate curve for each year. Note that HSGPA is positively associated with ACT-C score for all eight years. The slight flattening at the upper end of the curves shows a ceiling effect for conditional average HSGPA.

The vertical layering of the curves indicates grade inflation or deflation across years. This graph shows that the eight curves lie on top of each other with no definite pattern of annual grade change. Although no discernible evidence of systematic grade inflation can be identified, there are differences across years at different levels of ACT-C score. For example, there was greater variability in annual grade change at the lower and upper ends of the ACT-C score scale. This variation did not, however, demonstrate systematic inflation or deflation across years.

The general finding of no discernible pattern of grade inflation is in contrast to the findings of Woodruff and Ziomek (2004a). To explore the differences in results between the present and former study further, Figure 11.3 shows the change in HSGPA for selected ACT-C scores. This figure is based on public high school graduates between 1991 and 2011 who took the ACT test as part of National or State testing, tested during the eleventh or twelfth grade, and scored between a 14 and 31. This graph examines the period investigated in the Woodruff and Ziomek (2004b) study (1991–2003) as well as the present research (2004–2011). As this graph illustrates, from 1991 to 2001 there was an increase in conditional HSGPA for the selected ACT-C scores. After 2003, there was comparatively little change in the conditional HSGPA scores. This pattern held regardless of ACT-C score.

Summary. 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, school-level variation in conditional HSGPA change was evident across the eight years.

The results of this study provide both positive and concerning messages to high schools and postsecondary institutions. The fact that no evidence of grade inflation was found at the national level suggests that average HSGPA has stabilized, which may alleviate some concerns about possible validity decay of HSGPA for measuring students' preparedness for college or the workforce. This is not to say, however, that grade inflation and deflation do not exist. The significant variation across schools identified in this study is evidence that HSGPA inflation or deflation occurs at some high schools.

Table 11.5 Public High School Demographic Variables by Year

	2004	2005	2006	2007	2008	2009	2010	2011
Number of high schools	11,608	11,718	11,820	11,923	11,983	12,048	12,092	12,092
Average HSGPA	3.28	3.29	3.30	3.30	3.29	3.28	3.29	3.29
Average ACT Composite score	20.83	20.89	20.97	21.08	21.04	21.01	21.11	21.15
State-tested population percentage	44.05	44.35	44.13	46.18	48.19	49.82	52.09	54.62
Racial/ethnic minority percentage	27.75	27.81	27.99	28.19	28.33	28.44	28.49	28.46
American Indian	1.94	1.93	1.92	1.90	1.89	1.88	1.88	1.81
African American	13.61	13.63	13.71	13.77	13.82	13.85	13.86	13.81
Hispanic	12.53	12.59	12.69	12.84	12.93	13.02	13.05	13.01
Free/reduced-price lunch eligible percentage	39.74	39.68	39.68	39.63	39.60	39.58	39.55	39.49
Free lunch	31.68	31.81	31.81	31.77	31.74	31.73	31.70	31.64
Reduced-price lunch	8.04	8.04	8.04	8.03	8.01	8.01	8.00	7.93

Source: Zhang & Sanchez (2013)

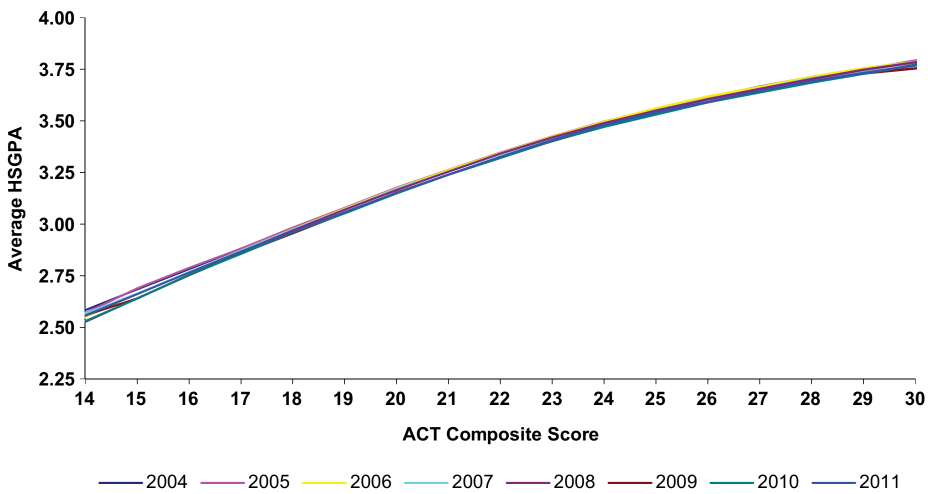


Figure 11.2 Plot of conditional HSGPA by ACT Composite score for the years of 2004 to 2011.

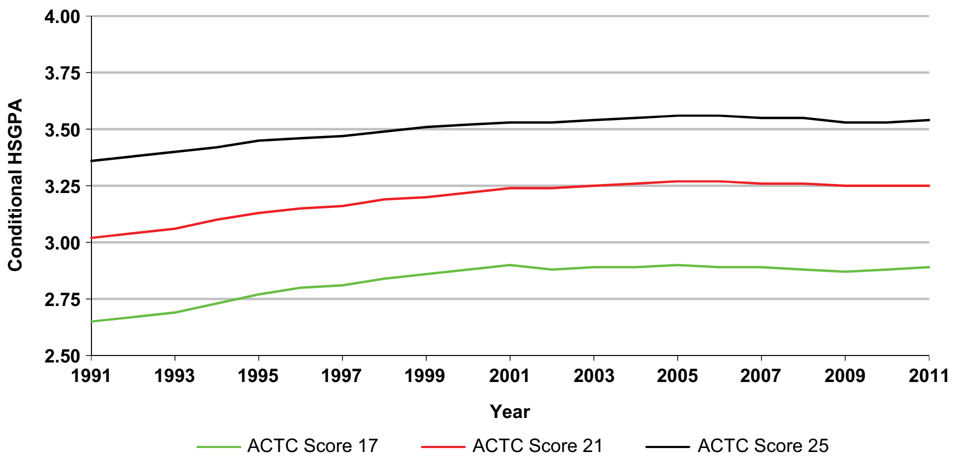


Figure 11.3 Plot of conditional HSGPA between 1991 and 2011 for selected ACT Composite scores.

Differential Grading Standards. Another study by Woodruff and Ziomek (2004b) was designed to assess how grading standards vary across high schools.

Data and method. The data included students who graduated from public high schools in the spring of 1998, 1999, 2000, 2001, and 2002 and took the ACT in the eleventh or twelfth grade. For each high school, the mean ACT Composite score was computed for each year. Only schools with at least 30 students were included. The schools were then divided into quintiles (i.e., five groups) based on the school means for each of the five years. The schools included in the analysis were those that remained in the 1st quintile group (bottom 20% of schools) and those that remained in the 5th quintile group (top 20% of schools) on the ACT Composite score for all five years. The number of schools in the 1st quintile group and the 5th quintile group were 664 and 573, respectively. Although the same schools were used for all five years, the graduating class of students in those schools changed from year to year. The hypothesis investigated was that schools in the 1st quintile group used more lenient grading standards than the schools in the 5th quintile group. HSGPA was regressed on ACT Composite score within each quintile group for each year. If the schools in the 1st and 5th quintile groups used the same grading standards, then the regression of HSGPA on ACT Composite score in the two quintiles should have had the same intercept and slope.

Results. Table 11.6 contains relevant statistics from the linear regression analyses. The results are similar for all five years. The two quintile groups have essentially equal slopes. Mean differences in grading practices between the two groups of schools equal the differences between their linear regression intercepts. The 1st quintile groups' mean leniency in grading ranged from a high of 0.19 in 1998 to a low of 0.12 in 2002; each was statistically significant ($p < .01$). In addition, the correlations between overall HSGPA and ACT Composite score were slightly higher for the 5th quintile group.

Figure 11.4 displays the regression lines estimating the linear relationship between overall HSGPA and ACT Composite score in 2000 for the 1st and 5th quintile groups (denoted Q1 and Q5, respectively). From the figure, it is clear that for students with the same ACT Composite score, the 1st quintile group had a higher mean overall HSGPA than the 5th quintile group.

Summary. The results of this study imply that grades are more of a relative standard in that they can vary from school to school. This study also 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 predictions of college readiness.

Table 11.6 Coefficients for the HSGPA on ACT Score Regressions for the First and Fifth Quintile in Each of the Five Years

Year	Quintile	N	Correlation	Slope	Intercept	Difference between intercepts
1998	Q1	53,939	.48	0.076	1.60	0.19
	Q5	96,586	.60	0.076	1.41	
1999	Q1	55,013	.49	0.077	1.60	0.16
	Q5	94,235	.60	0.076	1.44	
2000	Q1	59,434	.48	0.075	1.63	0.14
	Q5	101,833	.59	0.074	1.49	
2001	Q1	56,668	.47	0.075	1.66	0.14
	Q5	98,136	.59	0.073	1.52	
2002	Q1	52,997	.47	0.075	1.68	0.12
	Q5	86,536	.59	0.073	1.56	

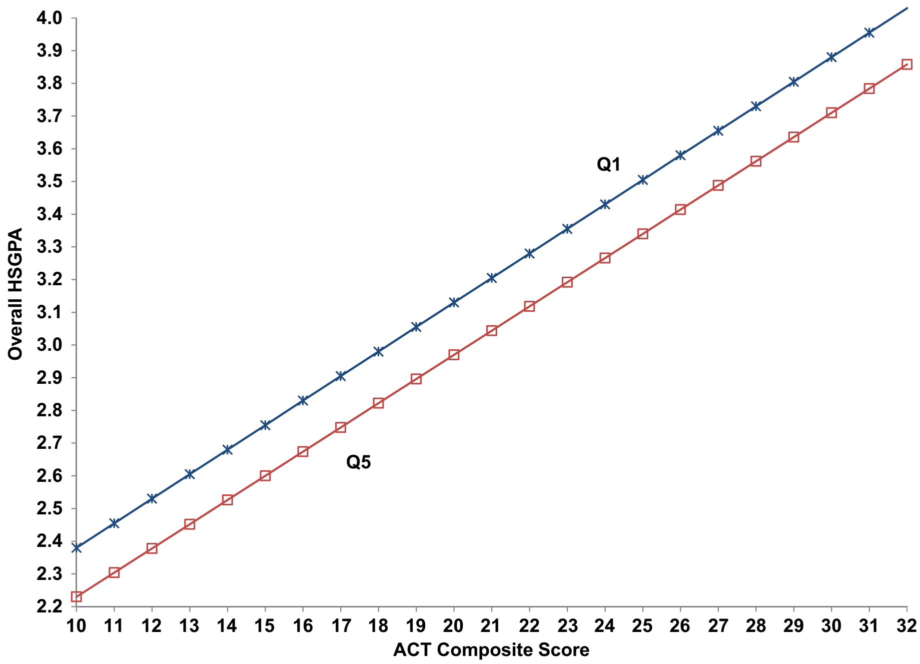


Figure 11.4 Plot of the year 2000 linear regressions of overall HSGPA on ACT Composite score.

11.2.4 Statistical Relationships between ACT Benchmark Attainment and High School Course Work and Grades

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 that for the ACT STEM and ELA Readiness Benchmarks, see Chapter 8.

A study by Ling and Radunzel (2017) examined how the high school course work taken and grades earned related to students' chances of meeting the ACT College Readiness Benchmarks in each of the four subject areas, after accounting for other student and school characteristics. The results of this study are described in this section.

Data and method. The study sample consisted of 6,440 high school seniors from 4,541 schools who took the ACT, which reflected a response rate of 12% of the sample invited to complete a supplemental online questionnaire about their high school experience, study and work habits, parental involvement, educational and occupational plans and goals, and college courses taken and/or college credits earned in high school (see McNeish et al., 2015 or section 11.2.2 for more details about the study sample). At the time they registered to take the ACT, students provided other information such as high school course work taken and grades earned. Students' readiness for college course work in a subject area was defined by whether the relevant ACT College Readiness Benchmark had been met.

A blockwise logistic regression model with cluster-robust standard errors was used to predict ACT Benchmark attainment from the student and school characteristics. Cluster-robust standard errors were used to account for students being sparsely clustered within high schools. A separate regression model was developed for each Benchmark. Candidate predictor variables were placed into the following five different blocks based on the nature of the variables: high school grades earned, courses taken, advanced and/or college-level course work taken in high school, school characteristics, and other noncognitive characteristics. Once a predictor was included based on statistical significance, it was retained in the model regardless of whether the statistical significance changed after subsequent blocks were added. Weights were applied in the analyses so that the study data resembled that of all 2012–2013 ACT-tested seniors nationally on student demographics and achievement levels.

The course work predictors included course sequence patterns in mathematics and science, individual courses in social studies (American Government, Geography, Economics, Other History, and Psychology), four separate indicators for whether advanced, honors, or dual-enrollment courses had been taken in a subject area, and the number of college course credits earned in high school. Grade-specific English courses were not included in the models because of the limited variability in students' course taking in this subject area.

Predictors were evaluated using a statistical significance level of .01. The adjusted odds ratio (OR) was used to describe the strength of the predictor-Benchmark attainment relationship. In comparison to a reference group, an OR greater than 1.0 indicates that students in the subgroup of interest are generally more likely to meet the Benchmark, whereas an OR less than 1.0 indicates that they are less likely to do so. For more details on the data and methods, see the full report (Ling & Radunzel, 2017).

Results. In this study, the weighted percentage of students meeting each of the ACT College Readiness Benchmarks was 67% in English, 46% in reading, 45% in mathematics, and 37% in science. Based on the Nagelkerke-R^2, the percentage of variance explained by the multiple-predictor models ranged from 39% (reading) to 55% (mathematics). Moreover, the multiple-predictor models correctly classified Benchmark attainment for 75% (reading) to 80% (English and mathematics) of the students, which represents a 19% (English) to 108% (science) increase over chance.

HSGPA was a strong predictor of Benchmark attainment in each of the subject areas; the adjusted OR associated with a one-unit change in HSGPA ranged from 2.9 in reading and science to 4.4 in mathematics (Table 11.7). HSGPA alone accounted for 20% (reading) to 30% (mathematics) of the variance in ACT Benchmark attainment.

Taking higher-level mathematics courses in high school was predicted to increase students' chances of meeting the Benchmarks in every subject area, while taking higher-level science course work was primarily associated with meeting the ACT Benchmark in mathematics (Table 11.7). For example, compared to students who took Algebra 1, Geometry, and Algebra 2, the odds of meeting the ACT Benchmark in mathematics was 1.7 times greater for students who also took either Trigonometry or another advanced mathematics course beyond Algebra 2, and 4.5 to 5.0 times greater for students who took a mathematics course sequence that included Calculus. Additionally, students who took accelerated, advanced, honors, and dual-enrollment course work in high school were more likely to meet the ACT Benchmarks. For example, the odds of meeting the ACT Benchmarks in English and reading were 1.6 to 1.7 times greater for students who took advanced, honors, and/or dual-enrollment courses in English compared to those who did not. Students expecting to earn college credits in high school were more likely than those expecting to earn zero college credits to meet the ACT Benchmarks in mathematics and science (adjusted OR = 1.1 to 1.4 for 1 to 6 credits and 1.3 for 7 or more credits). The course work taken in high school accounted for between 7% (in reading) and 16% (in mathematics) of additional variance. A more detailed description of the study results, including results for the other student and school characteristics in the models, is provided in the full report (Ling & Radunzel, 2017).

Summary. The study findings indicate that students who take rigorous courses in high school and earn good grades are more likely to meet the ACT Benchmarks and therefore are 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 are more strongly associated with preparation for postsecondary-level work than others. Each incremental college-preparatory course taken, particularly in mathematics and science (e.g., Trigonometry beyond Algebra 2, Physics beyond Chemistry), added to readiness more than did the number of courses 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 on the ACT CGIS, which does not provide more detailed information on the courses taken, especially in English.

Table 11.7 Adjusted ORs of ACT Benchmark Attainment

Predictor	English	Math	Reading	Science
HSGPA[a]	3.35	4.44	2.94	2.91
Mathematics course sequence				
Algebra 1, Algebra 2, and Geometry (referent)				
Algebra 1, Algebra 2, Geometry & Other Advanced Math	1.38	1.70	1.19**	1.47
Algebra 1, Algebra 2, Geometry & Trig	1.28**	1.69	1.14**	1.42
Algebra 1, Algebra 2, Geometry, Trig & Other Advanced Math	2.07	3.09	1.58	2.21
Algebra 1, Algebra 2, Geometry, Trig & Calculus	1.73	4.52	2.05	3.02
Algebra 1, Algebra 2, Geometry, Other Advanced Math, Trig & Calculus	1.92	5.00	2.07	3.34
Science course sequence				
Biology (referent)				
Biology & Chemistry	1.28**	1.73	—	1.12*
Biology, Chemistry & Physics	1.29*	2.31	—	1.42**
Individual social studies courses				
Psychology	1.17**	—	1.07**	—
Other History[b]	1.10**	—	1.15**	—
Advanced high school course work[c]				
Advanced English (taken/not taken)	1.63	0.88**	1.73	—
Advanced mathematics (taken/not taken)	1.33	2.13	—	1.66
Advanced social studies (taken/not taken)	1.34	1.39	1.62	—
Advanced science (taken/not taken)	1.29*	1.48	—	1.60
College credits earned in high school				
0 (referent)				
1 to 6	—	1.39	—	1.09*
7 or more	—	1.32	—	1.28

* indicates that the indicator was not statistically significant at the .01 level upon entry but was retained as part of a factor.

** indicates that the predictor was statistically significant upon entry but was no longer significant in the final model.

[a] Average of course grades in 23 core courses in English, mathematics, natural sciences, and social studies. This variable was grand-mean centered at 3.31.

[b] Other History course besides American History and World History.

[c] Advanced course work includes any accelerated, advanced, honors, and dual-enrollment courses taken in the subject area by the student while in high school.

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

ACT research has shown that taking rigorous, college-preparatory mathematics courses is associated with higher ACT mathematics and Composite scores (e.g., ACT, 2016b; 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 course work 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).

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 recent study (Allen, 2015b) examined the extent to which ACT scores are predicted by measures of achievement in specific core high school courses, controlling for pre-high school academic achievement.

Data and method. In this study, ACT Explore® scores served as measures of pre-high school educational achievement, and ACT QualityCore® scores measured high school course achievement. The ACT is based on the philosophy that the tests should measure the academic skills necessary for education after high school and the content of the tests should be related to major curriculum areas. The ACT focuses on the knowledge and skills attained through the cumulative effects of school experience. ACT Explore measured the knowledge and skills that are usually attained by Grade 8. ACT QualityCore included end-of-course assessments that measured performance against empirically-derived course standards. Students who took the ACT Explore tests in Grade 8, ACT QualityCore end-of-course exams in Grades 9, 10, or 11, and the ACT in Grades 11 or 12 were included in the study.

For each subject area of the ACT, same-subject ACT QualityCore end-of-course exams were used in the analysis. For English, ACT QualityCore scores from English 9, English 10, and English 11 were used; for mathematics, ACT QualityCore scores from Algebra 1, Geometry, and Algebra 2 were used; for reading, ACT QualityCore scores from US History were used; and for science, ACT QualityCore scores from Biology and Chemistry were used. Scores from other ACT QualityCore courses (English 12, Precalculus, and Physics) were not used because few students took the end-of-course exams for these courses, or a majority took them after taking the ACT. For students who took the ACT more than once, their last set of scores was used for analysis. ACT QualityCore scores were only used if the student took the ACT QualityCore course before or concurrently with the ACT (e.g., students who took an ACT QualityCore end-of-course exam and the ACT in spring of Grade 11 were included). The students included in the analyses were scheduled to complete high school between 2011 and 2016. For details on the sample used for each analysis, see the original study (Allen, 2015b).

Multiple linear regression was used to relate the measures of pre-high school educational achievement (ACT Explore scores) and high school course achievement (ACT QualityCore scores) to ACT scores. Results include regression coefficients, standard errors, *p*-values, and standardized beta weights. The regression coefficients represent expected ACT score changes for each one-point increase in the predictor while holding the other predictors constant. The standardized beta weights estimate how many standard deviations the mean ACT score changes for each one-standard-deviation increase in the predictor and allow for comparisons of the strengths of the relationships across predictors. If mastery of high school course content is positively related to ACT scores, the regression coefficients for the ACT QualityCore scores should be positive and statistically significant (i.e., *p*-value less than 0.05).

Results. End-of-course achievement in English 9, English 10, and English 11 was predictive of performance on the ACT English test, after controlling for pre-high school academic achievement (Table 11.8). That is, performance on the ACT English test is related to mastery of English courses in high school. With the exception of the ACT Explore reading score, all measures were statistically significant predictors of the ACT English score. The strongest predictive weights were observed for ACT QualityCore English 11 scores (beta = 0.290), Grade 8 ACT Explore English scores (beta = 0.269), ACT QualityCore English 10 scores (beta = 0.166), and ACT QualityCore English 9 scores (beta = 0.107).

End-of-course achievement in Algebra 1, Geometry, and Algebra 2 was predictive of performance on the ACT mathematics test, after controlling for pre-high school academic achievement (Table 11.9), indicating that performance on the ACT mathematics test is related to mastery of core mathematics courses in high school. All measures of pre-high school and end-of-course achievement were significant predictors of the ACT mathematics score. The strongest predictive weights were observed for ACT QualityCore Geometry scores (beta = 0.236), ACT QualityCore Algebra 2 scores (beta = 0.227), Grade 8 ACT Explore mathematics scores (beta = 0.209), and ACT QualityCore Algebra 1 scores (beta = 0.161). Level of achievement in courses with the closest time proximity to the ACT (e.g., Algebra 2, Geometry, and English 11) was more predictive.

End-of-course achievement in US History was predictive of performance on the ACT reading test, after controlling for pre-high school academic achievement (Table 11.10). The strongest predictive weights were observed for ACT QualityCore US History scores (beta = 0.347), Grade 8 ACT Explore English scores (beta = 0.252), and Grade 8 ACT Explore reading scores (beta = 0.220).

End-of-course achievement in Biology and Chemistry was predictive of performance on the ACT science test, after controlling for pre-high school academic achievement (Table 11.11). As was the case with the other ACT multiple-choice tests, performance on the ACT science test is related to mastery of science courses in high school. The strongest predictive weights were observed for ACT QualityCore Chemistry scores (beta = 0.267), ACT QualityCore Biology scores (beta = 0.229), Grade 8 ACT Explore mathematics scores (beta = 0.150), and Grade 8 ACT Explore science scores (beta = 0.131).

Summary. The results of the analyses support the proposition that performance on the ACT is related to 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 the use of ACT scores as measures of educational achievement.

The predictive weights of the course achievement measures 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 are strong predictors of ACT scores, results show that achievement in core high school courses also has a strong relationship with ACT scores. Students who master core high school courses are more likely to demonstrate high academic growth during high school.

In comparison to the McNeish et al. (2015) study (discussed in section 11.2.2), the models in this study explained a larger percentage of the variation in ACT scores. Prior achievement and achievement in core high school courses predicted ACT scores better than high school course grades and courses taken, high school characteristics, noncognitive characteristics, SES, and demographic variables. This may be due to the standardized measures of prior achievement and achievement in core high school courses being more directly related to the outcome, which was also a standardized measure of academic achievement, relative to unstandardized variables such as high school course work and grades.

Table 11.8 Predicting ACT English Score

Predictor	Estimate	SE	*p*	Beta
ACT Explore English	0.403	0.020	<.001	0.269
ACT Explore Mathematics	0.153	0.020	<.001	0.089
ACT Explore Reading	0.035	0.020	.077	0.022
ACT Explore Science	0.078	0.024	.002	0.040
ACT QualityCore English 9	0.100	0.013	<.001	0.107
ACT QualityCore English 10	0.172	0.015	<.001	0.166
ACT QualityCore English 11	0.266	0.013	<.001	0.290

Note. N = 4,336, R^2 = 0.732

Table 11.9 Predicting ACT Mathematics Score

Predictor	Estimate	SE	*p*	Beta
ACT Explore English	0.102	0.013	<.001	0.090
ACT Explore Mathematics	0.290	0.016	<.001	0.209
ACT Explore Reading	0.044	0.013	<.001	0.037
ACT Explore Science	0.139	0.017	<.001	0.095
ACT QualityCore Algebra 1	0.162	0.010	<.001	0.161
ACT QualityCore Geometry	0.238	0.010	<.001	0.236
ACT QualityCore Algebra 2	0.231	0.010	<.001	0.227

Note. N = 5,732, R^2 = 0.690

Table 11.10 Predicting ACT Reading Score

Predictor	Estimate	SE	*p*	Beta
ACT Explore English	0.371	0.004	<.001	0.252
ACT Explore Mathematics	0.094	0.004	<.001	0.055
ACT Explore Reading	0.355	0.004	<.001	0.220
ACT Explore Science	0.141	0.005	<.001	0.073
ACT QualityCore US History	0.395	0.002	<.001	0.347

Note. N = 134,470, R^2 = 0.650

Table 11.11 Predicting ACT Science Score

Predictor	Estimate	SE	*p*	Beta
ACT Explore English	0.121	0.014	<.001	0.097
ACT Explore Mathematics	0.222	0.016	<.001	0.150
ACT Explore Reading	0.102	0.014	<.001	0.076
ACT Explore Science	0.214	0.019	<.001	0.131
ACT QualityCore Biology	0.199	0.009	<.001	0.229
ACT QualityCore Chemistry	0.223	0.009	<.001	0.267

Note. N = 7,573, R^2 = 0.624

11.2.6 Statistical Relationships between ACT Scores and Noncognitive Factors

ACT has conducted a good deal of research examining the relationship between ACT scores and other student characteristics or noncognitive factors. Table 11.12 provides information from the Academic Skills Study, which included a nationally representative sample of 1988 ACT-tested students (ACT, 1997), about the relationships among scores, grade level, and educational plans. In this table, the means for the college-bound group are higher than the means for the national group for all four test scores and the Composite score. This finding indicates that, as expected, high school students who plan to attend college earn higher ACT scores than the typical high school student. Similarly, states that adopt the ACT statewide observe a sizable dip in average ACT scores post-adoption (Allen, 2015a). Also, as expected, ACT scores are related to grade level, with examinees in higher grades earning higher test scores, on average. These results indicate that the ACT measures what is taught in the classroom and that exposure to more content as signified by more years of schooling is related to higher ACT scores. A more direct examination of the relationship between course taking patterns and ACT performance is provided in sections 11.2.2 and 11.2.5.

Table 11.12 Scale Score Means and Standard Deviations of ACT Tests by Grade Level for the 1988 Nationally Representative Sample

		Grade 10		Grade 11		Grade 12	
Test/Subtest	Number of items	Mean	SD	Mean	SD	Mean	SD
National							
English	75	15.27	4.42	16.24	4.82	17.18	5.30
Mathematics	60	15.68	2.86	16.63	3.58	17.44	4.45
Reading	40	14.94	5.33	16.17	5.89	17.18	6.42
Science	40	16.31	3.71	16.91	3.93	17.48	4.34
Composite		15.67	3.51	16.61	3.99	17.45	4.54
College-Bound							
English	75	15.86	4.44	16.91	4.84	18.01	5.27
Mathematics	60	16.00	2.92	17.05	3.67	18.00	4.56
Reading	40	15.59	5.38	16.92	5.93	18.01	6.47
Science	40	16.72	3.74	17.36	3.98	17.99	4.41
Composite		16.17	3.54	17.18	4.02	18.13	4.56

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 11.13, students with higher ACT scores tended to describe their high school curriculum as college-preparatory in nature (r = .31 to .35) and aspire to higher educational levels (r = .32 to .36). Those who reported not needing help with their reading (r = .02 to .17), study skills (r = .07 to .09), and math skills (r = .10 to .26) tended to have higher ACT scores.

Table 11.13 Correlations among ACT Scores and Background Characteristics

	Educational needs, plans, and courses taken				
			Do not need help in[c]		
ACT Score	College-preparatory curriculum[a]	Educational plans[b]	Reading	Study skills	Mathematics
English	.34	.34	.14	.08	.13
Mathematics	.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 (voc-tech program, associate degree, bachelor degree, 1 to 2 years of grad 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 course work taken and grades earned, school characteristics, demographic characteristics, and ACT scores, with an emphasis on noncognitive measures related to students' academic goals, behaviors, perceptions, and parental involvement. The remainder of this section describes the results from this study (see also section 11.2.2 for more details about the sample).

Data. In this study, measures of noncognitive characteristics were obtained from 6,440 high school seniors through their responses to SPS items provided at the time they registered to take the ACT, as well as through their responses to an online questionnaire sent to them after they took the ACT. The online questionnaire asked students about their high school experience, study and work habits, parental involvement, educational and occupational plans and goals, and college courses taken and/or college credits earned in high school. All item response options on the online questionnaire were discrete and consisted of 5- or 6-point Likert-type items eliciting respondents' general level of agreement or frequency of partaking in a particular behavior or action.

Measures of students' academic goals and behavior included educational aspirations; an academic commitment component (composed of four questionnaire items related to how engaged a student is with academics); hours spent studying per week outside of class; hours spent working for pay per week; and whether the students had taken the ACT prior to their senior year. Measures related to students' perceptions included self-reported indicators for needing help in improving their skills in a variety of subject areas, a perception of education component (composed of three items assessing the value that students place on education), a self-assessment of whether their high school curriculum was a college-preparatory one, and an evaluation of how often their school challenged them to perform to the best of their ability. Parental involvement-related items asked about the frequency with which parents or guardians check that they have completed their assignments and the level of involvement of their parents or guardians in their post-high school plans.

Results. The block of noncognitive variables shown in Table 11.3 accounted for 4% to 7% of the variance in ACT scores, over and above the variance accounted for by high school grades and course work taken and school characteristics (see also Figure 11.1). The unstandardized regression coefficients reported in Table 11.3 can be interpreted as the expected change (increase or decrease) in ACT scores associated with a one-unit change in an independent variable, holding other variables in the model constant.

Several aspects of students' educational goals and values were positively related to performance on the ACT. Specifically, ACT scores were generally higher on average for students who (1) reported plans to pursue a postbaccalaureate degree; (2) described their high school course work as a college-preparatory curriculum; (3) took the ACT prior to their senior year; and (4) had higher perceptions on the value of education. Students who indicated that they needed help with their educational or occupational plans were predicted to score higher on the ACT English test (by 0.4 points).

The remaining significant noncognitive predictors were negatively related to performance on the ACT. For instance, students indicating that they need help with certain academic skills were predicted to have lower ACT scores, on average. Students reporting that they need help on reading speed and comprehension scored lower on all tests except mathematics (by 0.9 to 2.4 points), and students reporting that they need help with their math skills were predicted to have lower ACT mathematics, science, and Composite scores (by 0.5 to 1.5 points). Additionally, ACT scores in all four subject areas were negatively related to the frequency at which students felt challenged by their high school course work and the frequency at which their parents checked their assignments. The following noncognitive characteristics did not enter the ACT score models: the academic commitment component, hours spent studying per week outside of class, hours spent working for pay per week, and the level of parental involvement in students' post-high school plans. The block of SES-related demographic characteristics that included annual family income, parents' level of education, and whether English was the primary language spoken in the home explained an additional 1% of the variance in ACT scores, over and above the other variables in the models (Figure 11.1 and Table 11.3).

Summary. Consistent with results from earlier studies by Noble et al. (1999a, 1999b), this study found the contribution of selected noncognitive variables to the explanation of ACT performance, relative to course work taken, grades, and school characteristics was small (i.e., less than 10%). However, students' noncognitive characteristics alone explained 29% of the variance in HSGPA (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 course work 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 course work and the grades they earn in those courses, which, in turn, are strongly related to ACT scores.

11.2.7 Understanding Subgroup 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 course work, 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 may be largely attributable to differential academic preparation across student demographic groups.

Table 11.14 shows, by racial/ethnic group, the percentage of 2015–2016 ACT-tested high school graduates who completed a college-preparatory core curriculum, the percentage who had HSGPAs of 3.0 or higher, and the average ACT Composite scores for core completers and noncompleters. Students for whom the core indicator was missing were excluded from the calculations. The results indicate that students who completed a core curriculum tended to have higher ACT Composite scores, regardless of their race/ethnicity. For these students, mean ACT Composite scores ranged from 17.8 (for African American/Black students) to 24.7 (for Asian students). For students who did not complete a core curriculum, mean ACT Composite scores ranged from 15.7 (for African American/Black students) to 22.1 (for Asian students).

Table 11.14 Descriptive Statistics for ACT Composite Scores by Racial/Ethnic Group, 2015–2016

Ethnic group	% with core or more[a]	% with HSGPA ≥ 3.0[b]	Average Composite score	
			Core or more	Less than core
Black/African American	64	51	17.8	15.7
American Indian/Alaska Native	57	56	18.9	16.3
White	73	76	23.2	20.0
Hispanic/Latino	69	64	19.5	17.3
Asian	78	88	24.7	22.1
Native Hawaiian/Other Pac. Isl.	61	62	20.1	16.6
Two or more races	70	69	21.9	19.0

Note. [a] Students for whom the core indicator was missing were excluded from the calculations.
[b] Students for whom HSGPA data was missing were excluded from the calculations.

The extent to which ACT scores vary by gender has also been examined (Table 11.15). ACT Composite score averages were slightly higher for males than for females for most years; averages for both groups were relatively stable across years. The remainder of this section describes a study that examined the extent to which differential performance of various subgroups is potentially explained by factors such as academic achievement, courses taken, school characteristics, and noncognitive variables.

Table 11.15 Average ACT Scores by Gender, 2012–2016

			ACT score				
Gender	Reference year	N	English	Mathematics	Reading	Science	Composite
Female	2011–12	900,625	20.9	20.6	21.4	20.5	21.0
	2012–13	954,919	20.6	20.5	21.4	20.4	20.9
	2013–14	977,127	20.7	20.5	21.5	20.5	20.9
	2014–15	1,013,212	20.8	20.4	21.6	20.6	21.0
	2015–16	1,074,049	20.6	20.3	21.6	20.6	20.9
Male	2011–12	761,554	20.2	21.7	21.2	21.4	21.2
	2012–13	835,431	19.8	21.4	20.9	21.2	20.9
	2013–14	856,651	20.0	21.4	21.1	21.2	21.1
	2014–15	895,775	20.0	21.3	21.2	21.3	21.1
	2015–16	971,383	19.8	21.0	21.0	21.1	20.9

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 (see Section 11.2.2 for more details about the study sample). This study investigated the extent to which differential ACT performance among student demographic groups can be explained by high school grades, courses taken, school characteristics, and noncognitive characteristics related to students' academic goals, behaviors, perceptions, and parental involvement.

In the study, about 44% to 61% of the variability in ACT scores was attributable to specific course work taken and grades earned in high school; school characteristics; noncognitive characteristics related to students' academic goals, behaviors, and perceptions; parental involvement; and student demographics including race/ethnicity, annual family income, highest parental educational level, and gender (see Figure 11.1; variables were entered into each model in the order specified in the figure legend). About 28% to 46% of the variability in ACT scores was attributable to specific course work taken and HSGPA. As illustrated earlier in Figure 11.1, HSGPA explained substantial variance in ACT scores. School characteristics explained an additional 7% to 9% of the variability, and noncognitive characteristics explained an additional 4% to 7%. No more than 4% of additional variability was explained by student demographic characteristics (Table 11.16; 1% for the SES-related demographic characteristics and 1% to 3% for gender and race/ethnicity combined).

Table 11.16 presents the unstandardized regression coefficients for the student demographic comparisons (i.e., SES-related characteristics, gender, and race/ethnicity), after adjusting for the other cognitive and noncognitive variables and school characteristics earlier shown in Table 11.3. Statistically controlling for these other variables resulted in substantial reductions in mean ACT score differences between racial/ethnic, family income, and parental education groups. Comparisons between adjusted and unadjusted means by family income, race/ethnicity, and gender are presented next.

Table 11.16 Weighted Regression Statistics for Student Demographic Characteristics From ACT Score Models

	ACT score				
Predictor	English	Mathematics	Reading	Science	Composite
SES-Related Demographics					
English spoken at home—Yes vs. No	0.99	—	0.91	0.68	0.70
Annual Family Income[a]					
Middle vs. Low	0.37††	0.16**	—	0.22**	0.24**
High vs. Low	0.61	0.46	—	0.26**	0.39
Highest Parental Education Level					
Some College vs. No College	0.56	0.15*	0.54	0.21*	0.36
Bachelor's Degree vs. No College	0.91	0.35**	0.89	0.34**	0.61
Graduate Degree vs. No College	1.14	0.35**	1.11	0.44**	0.73
Increase in Total R^2 for SES-Related Demographics	.01	< .01	.01	.01	.01
Gender and Race/Ethnicity					
Gender—Female vs. Male	—	–1.14	—	–1.19	–0.64
Race/Ethnicity					
African American vs. White	–2.28	–1.67	–2.13	–2.07	–2.04
Hispanic vs. White	–1.98	–1.11	–1.66	–1.41	–1.53
Asian American vs. White	–1.24	0.85	–1.43	–0.58	–0.57
Other vs. White	–0.71	–0.28*	–0.32*	–0.43*	–0.44
Increase in Total R^2 for Gender and Race/Ethnicity	.02	.03	.01	.03	.02

Note. Regression coefficients for all student demographic variables were statistically significant ($p < .01$) unless denoted otherwise. Adjustment was made for cognitive, school-level, and noncognitive variables shown in Table 11.3.

† indicates a *p*-value between 0.010 and 0.015 upon entry to final model.

†† indicates a *p*-value between 0.010 and 0.015 in the final model.

* indicates that the indicator was not statistically significant upon entry but was retained as part of a predictor.

** indicates that the predictor was statistically significant upon entry but was no longer significant in the final model.

[a] The three categories for annual family income included < $36,000 (low), $36,000 to $80,000 (middle), and > $80,000 (high).

Income. For annual family income, unadjusted mean differences in ACT scores ranged between 2.0 (mathematics) and 3.1 (English) points between middle- and low-income students and from 3.7 (science) to 5.3 (English) points between high- and low-income students. After accounting for other student and school variables, the mean differences were reduced by 87% to 95% (Figure 11.5). For example, the unadjusted mean difference in average ACT reading scores between high- and low-income students was reduced from 4.3 points to 0.2 points. Differences in mean ACT scores among parental education levels were reduced by at least 74% when other student and school characteristics were taken into account (see McNeish et al., 2015, for more details).

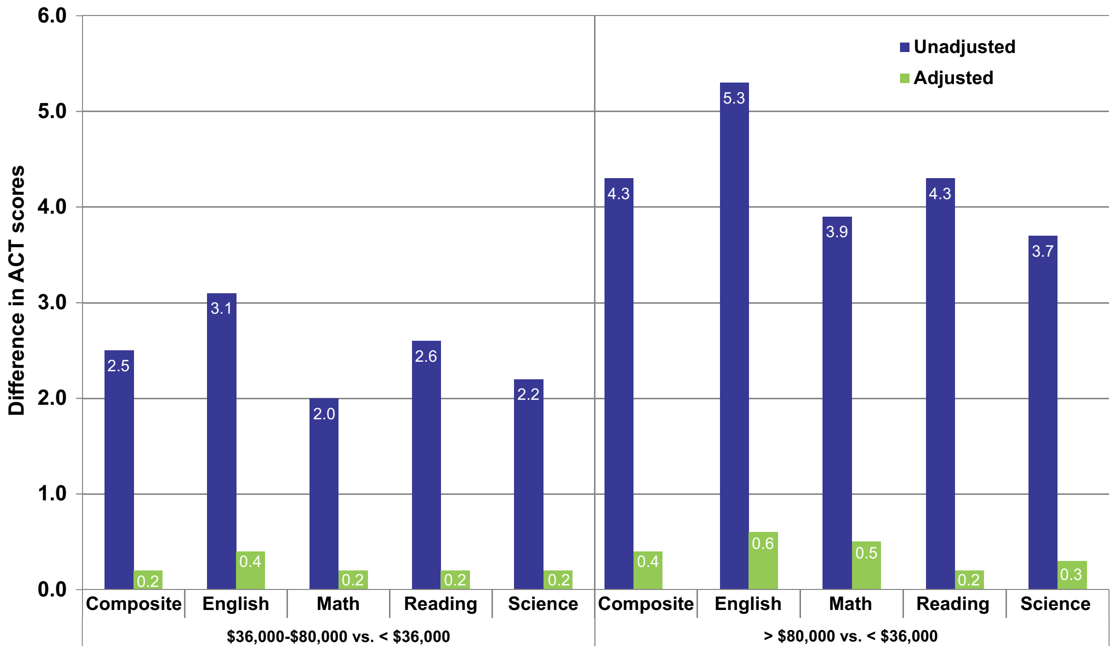


Figure 11.5 Unadjusted and adjusted mean differences in ACT scores by family income.

Race/Ethnicity. For race/ethnicity, unadjusted mean differences in ACT scores ranged from 4.2 points (mathematics) to 5.6 points (English) between White and African American students and from 2.7 points (mathematics) to 4.9 points (English) between White and Hispanic students. After adjusting for the other variables, mean differences were reduced by nearly 60% and ranged from 1.7 (mathematics) to 2.3 (English) between White and African American students and from 1.1 (mathematics) to 2.0 (English) points between White and Hispanic students.

Gender. For gender, differences in mean ACT mathematics, science, and Composite scores persisted, even after adjustment for other variables. In English and reading, adjusted mean scores did not significantly differ between male and female students. However, it should be noted that inferences about aggregate achievement or readiness drawn on self-selected groups, such as college-bound students, could be misleading. For example, Ndum and Mattern (2016) found that gender differences on the ACT mathematics and science tests were at least twice as large when based on a self-selected group of students as compared to results based on all eleventh-grade students within a state. An explanation for the differences in mean ACT mathematics, science, and Composite scores persisting in the McNeish et al. (2015) study is that the sample included students who self-selected to take the ACT because they were planning to attend college.

Summary. Results from this study suggest that differential performance on the ACT among student demographic groups is largely attributable to differential academic performance. Specifically, after accounting for HSGPA, high school course work taken, school characteristics, and other noncognitive factors, SES and other demographic characteristics accounted for a small percentage of the variance in ACT scores (4% or below). 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. In subsequent analyses, when the school-level demographic factors were excluded, student-level racial/ethnic and income regression coefficients were only slightly higher, by at most 0.4 point, than those reported. 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).

11.2.8 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 and/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.

This section focuses on a recent series of three studies in which 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. Results revealed that score increases were slightly greater for students who prepared for the second test, especially those who did not also prepare for the first test. Moreover, examinees who spent more hours preparing tended to increase their scores more than those who spent fewer hours preparing. Finally, examinees who perceived that test preparation helped them build confidence, become familiar with the test, refresh memory of content areas, and understand subject matter tended to increase their scores more than those who did not perceive test preparation to be helpful.

Data and method. Subjects for this study took the ACT in spring 2012 and fall 2012 or in fall 2012 and spring 2013. To isolate the effect of test preparation, analyses focused on the association between score increases and preparation for the second test. An online survey was sent to a sample of 76,000 repeat testers. The survey included questions about test preparation at the time of taking the ACT. A total of 9,654 examinees responded to the survey (12.7% response rate). Of those, 4,866 engaged in test preparation for their second test, and 4,788 did not.

The survey asked test takers whether they participated in test preparation activities or used any test-preparation materials (other than regular classroom instruction and assignments) before taking the ACT for the first or the second time. Survey respondents indicated the number of hours they spent on a variety of test-preparation activities including ACT's free preparation booklet, ACT's web-based preparation program (ACT® Online Prep), other web-based preparation programs, *The Real ACT Prep Guide*, another test-preparation workbook, test-preparation courses or workshops offered by an educational institution, commercial test-preparation courses, working with a private tutor or consultant, working one-on-one with a high school teacher, and test-preparation software. Finally, survey respondents indicated their perceptions of the ways in which test preparation was helpful. Possibilities included building confidence, becoming familiar with the test, refreshing memory of content areas, and helping understand subject matter.

Results. The first study analyzed data from all 9,654 survey respondents to determine whether preparation for a first or second test was associated with average ACT Composite score increases (Schiel & Valiga, 2014a). A greater percentage of respondents prepared for the second test than the first test (50.4% vs. 40.7%), 32.4% of respondents prepared for both tests, and 41.3% prepared for neither. Test takers who did not prepare for the second test increased their average ACT Composite score by 0.8 points, and those who prepared for the second test increased their average score by 1.4 points (Figure 11.6). Further disaggregation of the data revealed that examinees who prepared for the second test only had the largest average score gain (1.7 points). Students who did not prepare for either test increased their score by 0.9 points, on average.

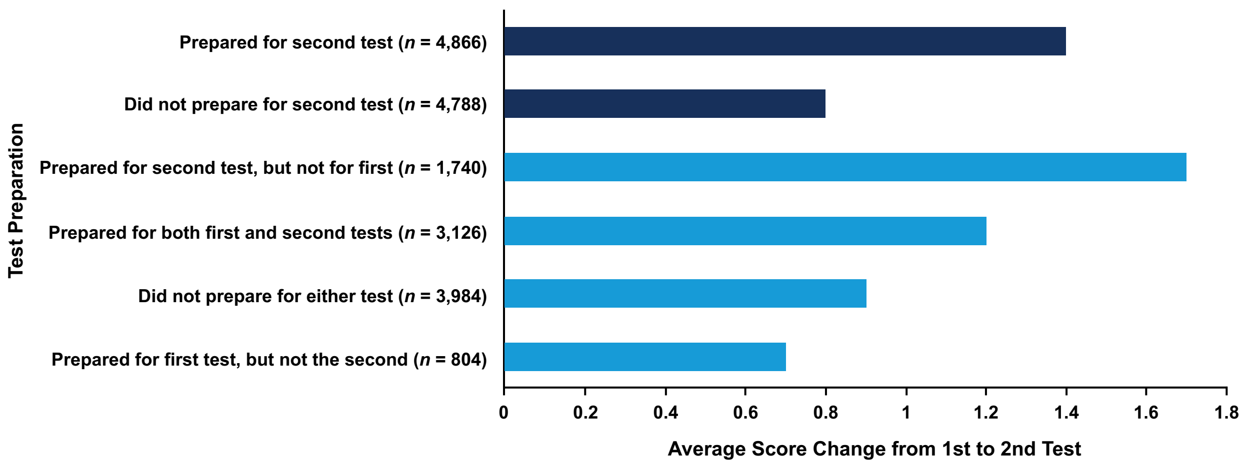


Figure 11.6 ACT Composite score changes of students who took the ACT on two occasions.

The second study revealed a positive association between hours of preparation for a second test and score increases (Schiel & Valiga, 2014b). That is, students who reported spending more time preparing to take the ACT for the second time had larger average ACT Composite score increases. Students who spent 3–6 hours preparing increased scores by an average of 1.0 point, students who spent 8–20 hours preparing increased an average of 1.4 points, and students who spent more than 20 hours preparing increased an average of 1.7 points (Figure 11.7). This positive association applies to all students who prepared, but the magnitude of average score increases differed depending on whether students also prepared for the first test. Specifically, average ACT Composite score increases were 0.4–0.6 points smaller for students who prepared for both tests compared to those who prepared for the second test only. The smallest average increase was 0.8 points, and this was observed for students who prepared for both tests but for a total of only 3–6 hours. Students who did not prepare for the first test but spent more than 20 hours preparing for the second test had the largest average score increase (2.1 points).

The third analysis estimated average ACT Composite score increases for students who perceived benefits from test preparation activities before their second test (Schiel & Valiga, 2014c). In response to the survey, students agreed or disagreed that test preparation imparted certain benefits: building confidence, becoming familiar with the test, refreshing memory of content areas, and understanding subject matter. On average, 87% of respondents agreed with each statement, suggesting that most test takers perceived benefits of test preparation. Average ACT Composite score increases were consistently higher for students who agreed with a given statement (Figure 11.8). For example, the average score increased by 1.5 points for students who reported that test preparation built their confidence, but the average score increased only 0.8 points for students who did not. Comparing test takers who agreed and disagreed that test preparation familiarized them with the test, the average score increased by 1.4 and 0.9 points, respectively. Test takers who agreed that test preparation refreshed their memory of content areas increased their scores by an average of 1.4 points, which was slightly higher than the 1.2 point average increase observed for those who disagreed. Test takers who indicated that test preparation helped them understand subject matter increased their average score by 1.5 points, whereas those who disagreed increased their average score by 1.1 points.

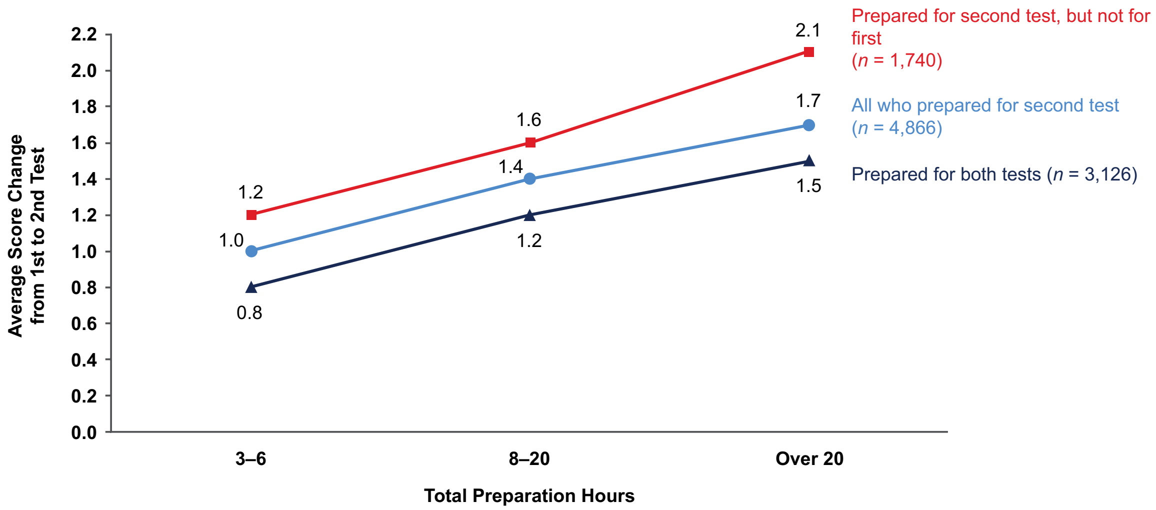


Figure 11.7 ACT Composite score changes of students with different durations of test preparation for the second test.[3]

[3] Student response options on the "hours spent on activity" items were in interval ranges. The estimated total preparation hours was derived for a student by first taking the midpoint of each interval response for the "hours spent on activity" items and then summing these individual estimated hours across the eleven "hours spent on activity" items. There were no students with 7 total preparation hours.

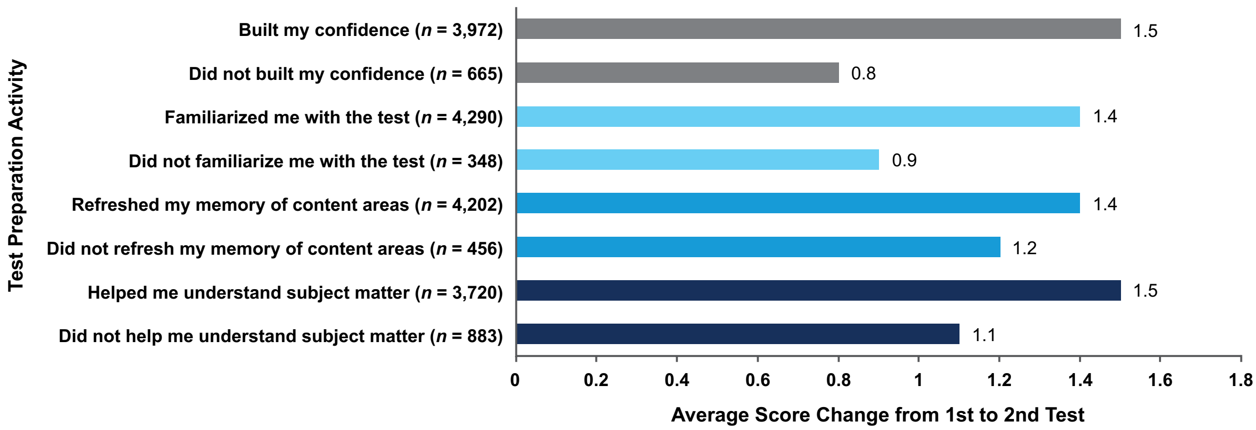


Figure 11.8 ACT Composite score changes of students with different perceptions of test preparation utility.

Summary. Schiel and Valiga (2014a, 2014b, 2014c) examined ACT Composite scores and test preparation survey responses for ACT examinees who tested in two consecutive semesters. Regardless of preparation, 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. 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 increase compared to those who did not. Further analyses 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. Most survey respondents 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.

In these studies, average increases were approximately 1–2 points, depending on the duration of preparation, and similar in magnitude to increases observed when students retake the test with no supplemental preparation (Andrews & Ziomek, 1998). These studies suggest that the average effect of preparation is small, but larger score increases (and decreases) could be expected for individual students. Although these studies cannot support causal conclusions about the effectiveness of test preparation activities, results 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.

11.2.9 Measuring 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, seventh- or eighth-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 11.9 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 sixth, seventh, or eighth grade in 2016 and sent their scores to a particular national talent search program.

Figure 11.9 shows that the cumulative distribution for the 2016 ACT-tested graduating class is shifted slightly to the right of the cumulative distribution for the talented search students who took the ACT in sixth, seventh, or eighth grade (average ACT Composite score: 20.8 vs. 18.6, respectively). This figure suggests that ACT scores can be used to measure the relatively greater educational development of academically talented students in sixth, seventh, and eighth grade. This is further substantiated by another study (Allen, 2016a) that compared ACT Composite scores earned by seventh graders from an earlier cohort to ACT Composite scores earned by the same students four or five years later; the average score gain was 9.4 points for these students. Average score gains were lowest for students who scored near the top of the ACT Composite scale as seventh graders.

A study by Schiel (1998) examined the academic benefits in high school of an intensive summer program for academically talented seventh 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).

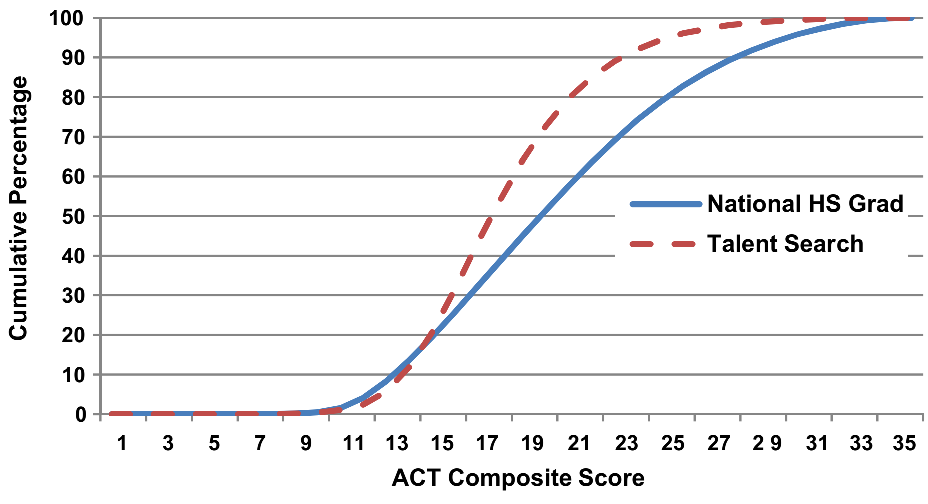


Figure 11.9 ACT Composite cumulative percentages for 2016 ACT-tested high school graduates and talent search sixth-, seventh-, and eighth-grade students.

11.3 Making 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), as well as decisions made by institutions.

One important aspect of success in college is academic achievement, and one critical determinant of academic achievement is academic preparation. 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 later go on to have satisfactory levels of achievement in subsequent years. Nevertheless, students whose academic difficulties in their first year cause them to leave college obviously 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, they do not measure all relevant academic skills. No practically feasible test is ever likely to do so. 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.

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.

11.3.1 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.

Traditional Validity Statistics. 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 is usually interpreted to mean that the relationship between college course work and test content is too weak for the test to be used for college admission.

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 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 these imperfect measures 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 measure were measured perfectly. Two recent studies have indicated that the estimated mean reliability of first-year GPA (FYGPA) to range between .75 and .87 (Beatty, Walmsley, Sackett, Kuncel, & Koch, 2015; Westrick, in press), which is lower than the reliability estimate of .94 for the ACT Composite score (see more on Reliability in Chapter 10). As an example, if the observed correlation between the ACT Composite score and FYGPA is .38, and the reliability estimate for FYGPA is .81, and the reliability of ACT scores is set to 1.0 (no correction), the validity coefficient for ACT Composite scores would increase from .38 to .42 ($.38/(\sqrt{(.81*1)}) = .42$).

Range restriction in variables also reduces the correlation between predictor and criterion measures, and it is an issue 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). The remainder of this section describes a recent validity study (Westrick, Le, Robbins, Radunzel, & Schmidt, 2015) that demonstrated the effects of range restriction.

Data. Data for the study included 189,612 ACT-tested students who enrolled in a four-year institution as first-time students entering in the fall term between 2000 and 2006, with each institution having between one and seven freshman cohorts. Fifty institutions that participated in various ACT research services or partnerships were represented. Available information also included the students' ACT scores, self-reported HSGPA, and self-reported parental annual income.

Method. For each institution, Pearson product–moment correlations were calculated between the following variables: ACT Composite scores, HSGPA, SES (self-reported parental income), and FYGPA. The correlations were then corrected for range restrictions in the three predictors (ACT, HSGPA, and SES) using the multivariate range restriction correction procedure introduced by Lawley (1943). Range restriction ratios on these predictors were computed for each institution based upon the standard deviations obtained from the institution and from the referent population (all ACT examinees between 1999 and 2006). The corrected correlations thus were estimates of correlations between the variables if they had been obtained in the referent population. The correlations were then meta-analytically combined across institutions (Hunter & Schmidt, 2004). To account for the increase in sampling error resulting from range restriction corrections, the Ree, Earles, and Teachout (1994) procedure was applied, and then effective sample sizes for each correlation were calculated. This allowed for a more accurate estimation of the variation across institutions due to sampling error.

Moderator analyses were conducted using three levels of institutional admission selectivity. The classifications were based on institutional self-reports of their admission selectivity. Highly selective and selective institutions were combined into a "high" selectivity category (k = 8); institutions with traditional admission selectivity policies were classified in the "mid" selectivity level (k = 29); and institutions with liberal and open admission selectivity policies were classified in the "low" selectivity level (k = 8). Four institutions did not report their admission selectivity, and they were excluded from the moderator analyses.

Results. Table 11.17 presents the observed mean correlations and the estimated mean population correlations between the original predictor variables and FYGPA. After corrections for range restriction, the estimated mean correlation between ACT scores and FYGPA was .51, and the estimated mean correlation between HSGPA and FYGPA was .58. The validity coefficients for ACT Composite score and HSGPA were somewhat variable across institutions, with 90% of the coefficients estimated to fall between .43 and .60 and between .49 and .68, respectively (as indicated by the 90% credibility intervals).

In contrast, after correcting for range restriction, the estimated mean correlation between SES and FYGPA was only .24 and did not vary across institutions. For all three predictor variables, the lower bounds of the credibility intervals exceeded zero, indicating that there were generally positive relationships between the predictors and the criterion.

Table 11.17 Meta-Analysis of Multi-Institution Data—Correlations with FYGPA, Overall Analyses

Predictors	*k*	*N*	Mean observed *r*	*SDr*	Estimated mean *ρ*	*SDρ*	*95% CI*	*90% CrI*
ACT Composite scores	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

Notes. *k* = number of institutional studies; *SDr* = standard deviation of observed correlations; *SDρ* = standard deviation of correlations corrected for artifacts; *CI* = confidence interval; *CrI* = credibility interval. Table adapted from Westrick et al., 2015.

Table 11.18 contains the results by institutional admission selectivity. Though the estimated mean correlations varied across the selectivity levels, the 95% confidence intervals overlapped. This would suggest that the differences in the estimated mean correlations were due to sampling error. As in the overall analyses, none of the 90% credibility intervals contained zero, indicating that the relationships between the three precollege predictors and FYGPA were positive in all cases.

Table 11.18 Meta-Analysis of Multi-Institution Data—Correlations with FYGPA, Moderator Analyses by Admission Selectivity

Predictors	Admission Selectivity	*k*	*N*	Mean observed *r*	*SDr*	Estimated mean *ρ*	*SDρ*	*95% CI*	*90% CrI*
ACT Composite scores	High	8	69,944	.36	.05	.54	.04	.51, .56	.48, .59
	Mid	29	80,750	.39	.08	.51	.05	.49, .53	.43, .54
	Low	8	11,357	.39	.11	.47	.11	.40, .55	.30, .65
HSGPA	High	8	62,145	.47	.03	.63	.06	.59, .67	.54, .72
	Mid	29	71,378	.48	.05	.57	.04	.55, .59	.50, .64
	Low	8	9,807	.45	.10	.50	.13	.41, .59	.29, .71
SES	High	8	55,176	.12	.01	.26	.00	.24, .27	.26, .26
	Mid	29	67,818	.12	.05	.24	.00	.23, .25	.24, .24
	Low	8	9,322	.11	.06	.23	.00	.20, .26	.23, .23

Notes. *k* = number of institutional studies; *SDr* = standard deviation of observed correlations; *SDρ* = standard deviation of correlations corrected for artifacts; *CI* = confidence interval; *CrI* = credibility interval. Table adapted from Westrick et al., 2015.

Summary. The estimated mean correlations of ACT Composite scores and HSGPA with FYGPA provide evidence supporting the use of these measures in making college admission decisions. The 90% credibility intervals indicate that the validities of ACT scores and HSGPA vary across institutions. That is, the strength of the relationship between the predictor measures and the criterion differs across institutions. Though the corrected correlations varied across institutions, the relationships were positive at all institutions, indicating that students entering college with higher ACT Composite scores and HSGPAs tended to earn higher grades in first-year courses than their peers with lower ACT Composite scores and HSGPAs earned.

Finally, the results of this study demonstrate the impact of range restriction on validity coefficients. The corrections for range restriction in the predictor measures increased the validity coefficients for all the predictors, with increases ranging between .05 and .18.

Decision-Based Statistics. The correlation coefficient is probably 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 course work. The correlation coefficient indicates 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, an alternative method that can be used for summarizing the results of predictive validity studies that utilizes logistic regression and decision-based statistics is described. Studies presented in subsequent subsections of this section (for making admission decisions) and the next section (for making course placement decisions) will demonstrate the use of this method.

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 results (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 to 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

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

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

$= 0$, 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 11.10. 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 Expression (1). Therefore, larger values of $\hat{b}$ in logistic regression curves correspond to steeper slopes and to better discrimination between students who will and will not succeed.

The estimated weights $\hat{a}$ and $\hat{b}$ in Expression (1) can be calculated by iterative least squares procedures. Given the previous discussion, the coefficient $\hat{b}$ should be both positive and statistically significant. 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

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

where $\hat{P}[W=1 \mid X=x]$ is Expression (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 $\hat{P}[W=1 \mid X=x]$ in Expression (1) are calculated from data for students who enrolled in the institution. The probability $\hat{P}[A]$ in Expression (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.

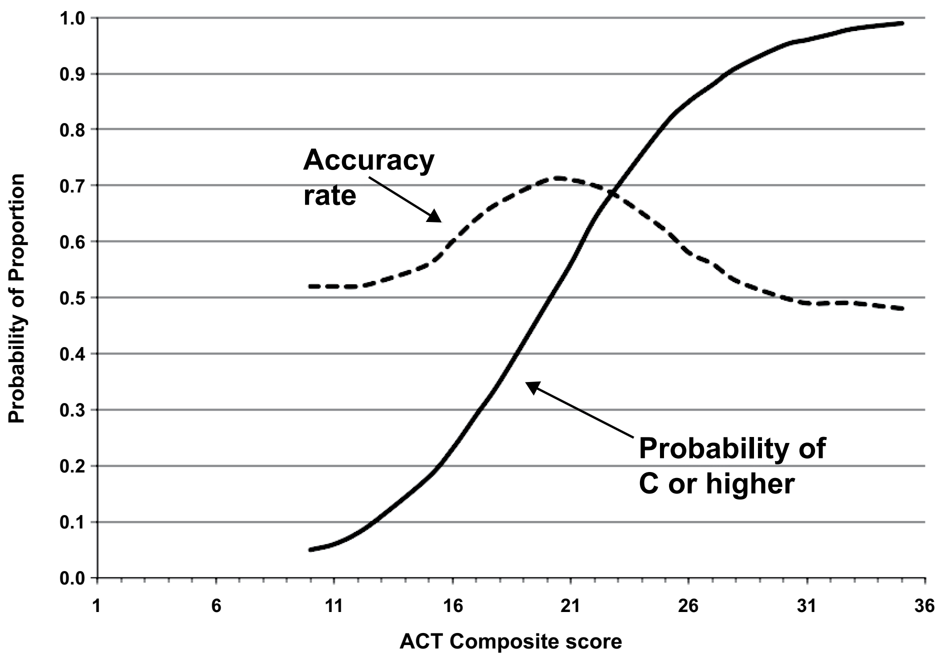


Figure 11.10 Probability of C or higher FYGPA and accuracy rate.

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 are shown later in this chapter that 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 11.10 as a proportion. Note that the maximum accuracy rate (.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 .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, shows 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 increases content coverage and, as a consequence, increases the accuracy of admission over that obtained by 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 goal of using that variable. Two common goals related to academic achievement are:

- Maximize academic success among enrolled students.
- Identify accurately those 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 whom 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 (Sawyer, 2010) using the decision-based statistics discussed in the previous section. Specifically, 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.

Data. The analyses were based on data from 192 four-year postsecondary institutions that used ACT scores in their admission procedures. The institutions provided outcome data either through their participation in ACT's predictive validity service or through participation in special research projects. The outcome data pertained to the following entering freshman class years: 2003, 2004, 2005, and 2006. The 192 institutions in the sample for this study had 483,451 nonenrolled score senders, in addition to their 120,338 enrolled students. Score senders (students who sent their ACT scores to particular institutions) were used as a proxy for applicants. For a more complete description of the study sample, see the full ACT Research Report (Sawyer, 2010).

Method. Academic success was defined jointly as retention through the first year and overall FYGPA. Students who completed the first year with a given FYGPA or higher were considered successful ($S = 1$); otherwise, they were considered unsuccessful ($S = 0$). The following four levels of success were considered:

- *S20*: Retention through first year and 2.0 or higher FYGPA (minimal success)
- *S30*: Retention through first year and 3.0 or higher FYGPA (typical level of success)
- *S35*: Retention through first year and 3.5 or higher FYGPA (high level of success)
- *S37*: Retention through first year and 3.7 or higher FYGPA (very high level of success)

Students who either dropped out or had a low FYGPA during their first year were unsuccessful. According to the study data, about 84% of students were at least minimally successful, about 52% were at least typically successful, about 27% were highly successful, and about 16% were very highly successful.

The conditional probabilities of success given the selection variable(s) were estimated using hierarchical logistic regression models. The models were constructed based on ACT Composite score (ACT-C), HSGPA, and ACT-C and HSGPA jointly. All the independent variables were centered about their respective grand means. The joint model included an interaction term between ACT-C and HSGPA.

From the estimated conditional probabilities of success, accuracy rates and success rates were calculated using the following cutoff proportions for each selection variable: .01, .10, .20, .30, .40, .50, .60, .70, .80, .85, .90, .95, and .99. These cutoff proportions correspond to increasing degrees of admission selectivity: the cutoff proportion .01 corresponds to admitting all but the bottom 1% of students, as ranked by their estimated probability of success; the cutoff proportion .99 corresponds to admitting only the top 1% of students.

Results. Figures 11.11 and 11.12 illustrate the typical probabilities of success calculated from the fixed-effect parameter estimates of HSGPA and ACT-C. In both graphs, the horizontal axis is scaled in terms of both the values of the selection variables and their associated cutoff proportions (or cumulative relative frequencies). A table of the parameter estimates is provided in the full ACT Research Report (Sawyer, 2010; see p. 29). In both of the single-variable models, the fixed effects for the HSGPA and ACT-C slopes are positive and statistically significant ($p < .001$). Moreover, the slope coefficients for HSGPA and ACT-C both increase with FYGPA success level. For example, the ACT-C slope coefficient is 0.16 for the 2.0 success level and 0.30 for the 3.7 level. Additionally, the variances of the HSGPA and ACT-C slope coefficients among institutions also increase with success level. This finding suggests that the strength of these variables' relationships with higher levels of FYGPA success varies more among institutions than does the strength of their relationships with lower levels of success.

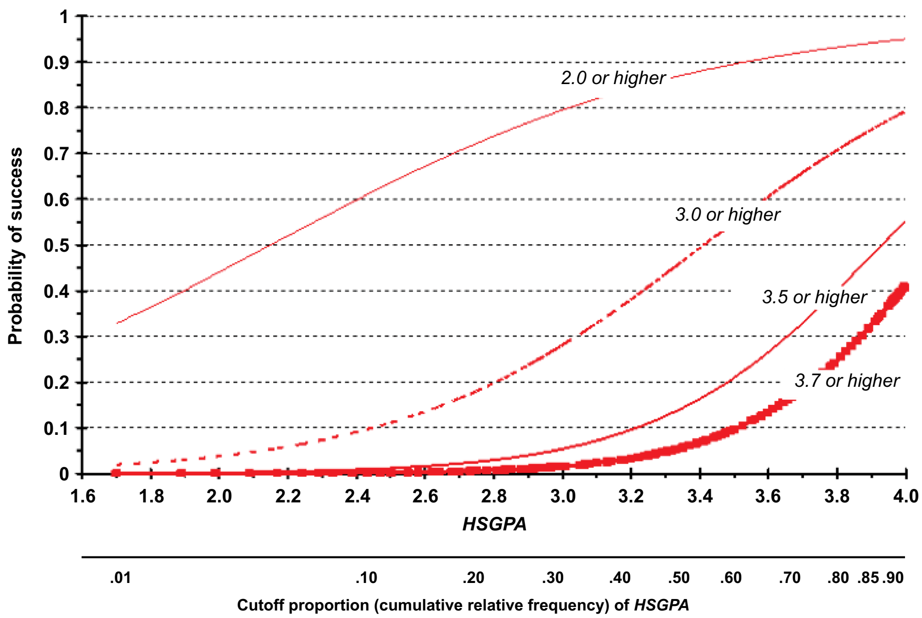


Figure 11.11 Probabilities of success associated with 2.0, 3.0, 3.5, and 3.7 or higher FYGPA and being retained through the first year, based on HSGPA.

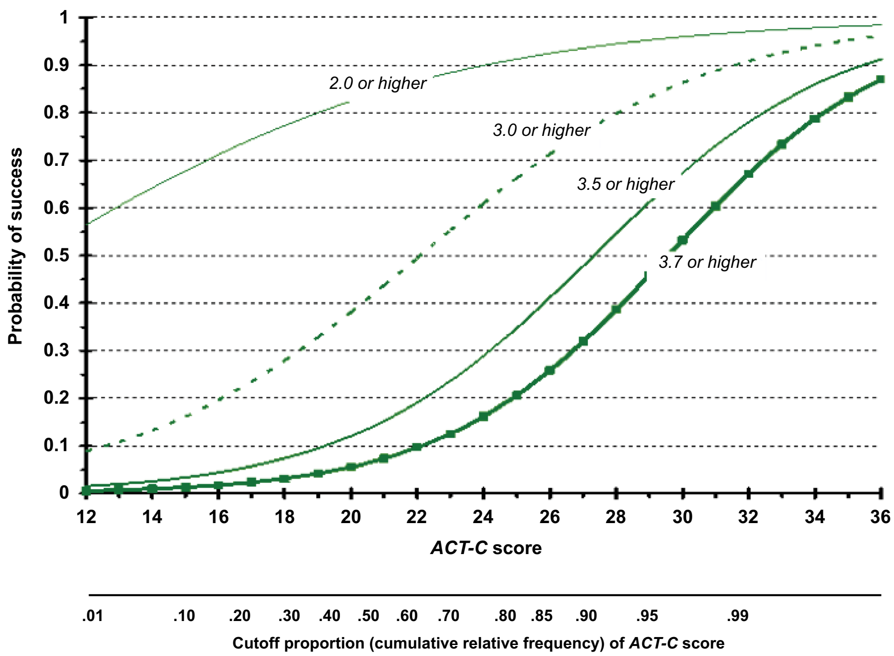


Figure 11.12 Probabilities of success associated with 2.0, 3.0, 3.5, and 3.7 or higher FYGPA and being retained through the first year, based on ACT Composite score.

Figure 11.13 shows the probability of earning a FYGPA of 3.0 or higher, given different values of HSGPA and ACT-C. The fixed effects for both the main effects and the interaction term between the two predictors were positive and statistically significant ($p < .001$). One interpretation of the interaction term is that HSGPA is more predictive among students with higher ACT-C scores than for students with lower ACT-C scores. That is, as ACT-C increases, the slope of the HSGPA probability-of-success curve increases.

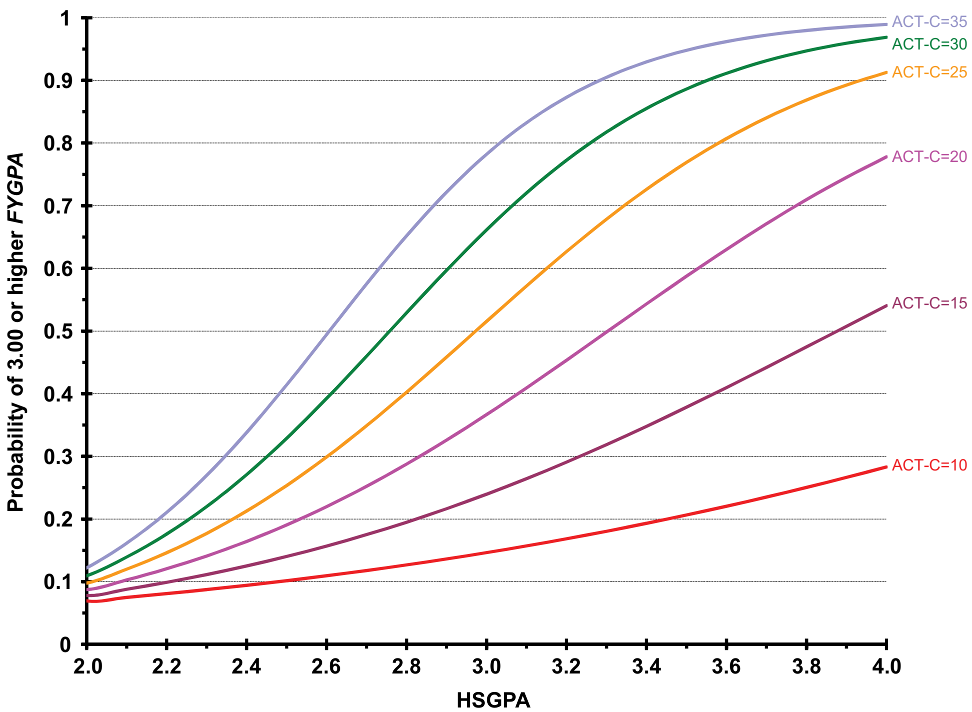


Figure 11.13 Probabilities of success associated with 3.0 or higher FYGPA and being retained through the first year, based on HSGPA and ACT Composite score.

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-C scores. Among students with a 4.0 HSGPA, students with an ACT-C score of 15 have a probability of .54 as compared to over a 0.95 probability for students with an ACT-C score of 30. Even for less extreme cases, the results illustrate that ACT-C score meaningfully discriminates among students with the same HSGPA.

Table 11.19 shows the median incremental success rates associated with the four success levels and the three sets of selection variables. The last row of the table shows a reference maximum, equal to one minus the median base success rate. Incremental success rates increase markedly with success level up to 3.5 but then decrease slightly at 3.7. For example, selection based on ACT-C results in a maximum incremental success rate of .14 for 2.0 or higher FYGPA, .45 for 3.0 or higher, .56 for 3.5 or higher, and .54 for 3.7 or higher. Note that HSGPA had higher incremental success rates than ACT-C at low to moderate cutoff proportions, but ACT-C did better than HSGPA at high cutoff proportions. Finally, at higher cutoff proportions, selection based on ACT-C and HSGPA jointly increased the incremental success rate over that for selection based on HSGPA or ACT-C alone.

Table 11.20 shows the median incremental accuracy rate with respect to null decisions of either admitting all applicants or denying admission to all applicants. The medians in each cell of this table are based on only those institutions at which the incremental accuracy rate is positive. For both the minimal level of success (2.0 or higher) and the very high level of success (3.7 or higher), the median incremental accuracy rate is often small (under .05). This result is a consequence of the relatively small reference maximums for these two success levels. As proportions of their reference maximums, however, the incremental accuracy rates are fairly large.

Table 11.19 Median Incremental Success Rate with Respect to Base Success Rate, by FYGPA Success Level, Cutoff Proportion, and Selection Variable (N = 192)

			Success level											
	Approx. value of		2.0			3.0			3.5			3.7		
Cutoff proportion	HS GPA	ACT-C	HS GPA	ACT-C	HS GPA & ACT-C	HS GPA	ACT-C	HS GPA & ACT-C	HS GPA	ACT-C	HS GPA & ACT-C	HS GPA	ACT-C	HS GPA & ACT-C
.01	1.7	12	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
.10	2.4	15	.03	.02	.03	.04	.03	.04	.02	.02	.02	.01	.01	.01
.20	2.7	17	.06	.03	.05	.08	.06	.07	.04	.04	.04	.03	.02	.03
.30	3.0	18	.07	.05	.07	.12	.09	.12	.07	.06	.07	.05	.04	.04
.40	3.2	19	.09	.06	.09	.15	.12	.16	.11	.08	.10	.07	.06	.06
.50	3.3	20–21	.10	.07	.10	.19	.16	.20	.14	.12	.14	.10	.08	.09
.60	3.5	22	.11	.08	.12	.23	.19	.26	.19	.15	.19	.13	.10	.13
.70	3.7	23	.12	.09	.13	.26	.24	.30	.23	.20	.25	.17	.14	.18
.80	3.8	25	.13	.11	.15	.30	.28	.37	.28	.26	.34	.21	.20	.25
.85	3.9	26	.13	.11	.15	.31	.31	.40	.31	.30	.39	.24	.24	.30
.90	3.95	27	.13	.12	.16	.33	.35	.43	.32	.36	.45	.26	.30	.37
.95	4.0	29	.13	.13	.17	.34	.39	.47	.34	.43	.53	.28	.39	.46
.99	4.0	31–32	.13	.14	.18	.34	.45	.51	.34	.56	.63	.29	.54	.61
Reference maximum				.20			.57			.80			.88	

Table 11.20 Median Incremental Accuracy Rate with Respect to Null Decisions among Institutions at Which It Is Positive, by FYGPA Success Level, Cutoff Proportion, and Selection Variable (N = 192)

			Success level											
			2.0			3.0			3.5			3.7		
	Approx. value of													
Cutoff proportion	HS GPA	ACT-C	HS GPA	ACT-C	HS GPA & ACT-C	HS GPA	ACT-C	HS GPA & ACT-C	HS GPA	ACT-C	HS GPA & ACT-C	HS GPA	ACT-C	HS GPA & ACT-C
.01	1.7	12	.01	.00	.00	.01	.01	.01	.01	.01	.01	.	.	.
.10	2.4	15	.03	.01	.02	.07	.04	.06	.06	.05	.08	.	.	.
.20	2.7	17	.05	.02	.04	.11	.07	.09	.15	.09	.17	.	.	.
.30	3.0	18	.08	.04	.06	.13	.09	.12	.23	.11	.25	.	.	.
.40	3.2	19	.08	.05	.07	.15	.09	.14	.09	.04	.08	.	.	.
.50	3.3	20–21	.07	.04	.07	.14	.10	.15	.07	.04	.08	.03	.03	.05
.60	3.5	22	.06	.04	.07	.12	.09	.14	.07	.05	.08	.05	.04	.08
.70	3.7	23	.07	.02	.07	.11	.09	.14	.05	.05	.09	.05	.03	.09
.80	3.8	25	.08	.06	.09	.09	.08	.11	.04	.04	.06	.02	.03	.05
.85	3.9	26	.04	.04	.07	.07	.07	.09	.03	.03	.05	.02	.03	.05
.90	3.95	27	.03	.03	.05	.05	.05	.07	.02	.02	.04	.01	.02	.04
.95	4.0	29	.02	.02	.03	.03	.03	.04	.01	.02	.03	.01	.01	.02
.99	4.0	31–32	.00	.01	.01	.01	.01	.01	.00	.01	.01	.00	.00	.01
Reference maximum				.20			.40			.20			.12	

Note. The "null decisions" are admitting every student or denying admission to every student. Empty cells indicate that no institutions had incremental accuracy with respect to null decisions.

For the 3.0 and the 3.5 success levels, median incremental accuracy rates are often larger than .05. For example, the joint HSGPA and ACT-C selection variable has maximum incremental accuracy near .15 for the 3.0 success level, and near .25 for the 3.5 success level. For all success levels, ACT-C has incremental accuracy beyond HSGPA at most institutions for cutoffs above HSGPA =3.3 or ACT-C=20 to 21 (see page 44 of the full ACT Research Report for the percentage of institutions with incremental accuracy with respect to null decisions by FYGPA success level, cutoff proportion, and selection variable).

Summary. The results from this study are 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. In some situations (for example, where an institution is interested in high levels of success), ACT Composite score is more useful. In most scenarios, using both high school grades and test scores jointly is better than using either by itself. In using both variables, it is important to take into account the HSGPA by ACT Composite score interaction effect, as well as the main effects.

In conclusion, 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 this study suggest that ACT Composite scores provide differentiation across levels of achievement in terms of students' probable success during their first year in college. For a more detailed description of these results, see the full ACT Research Report (Sawyer, 2010).

11.3.2 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 subgroups using ACT Composite scores (ACT-C) and HSGPA for making admission decisions. Subgroup characteristics included race/ethnicity, gender, and income. For each student subgroup, Sanchez examined the effect of using a total group cut score for ACT-C, HSGPA, or both on predicting first-year college grade point average (FYGPA).

Data. The data for the study included 259 two- and four-year institutions participating in ACT's Prediction Research Service or in special research projects (Sawyer, 2013a). The data consisted of more than 137,000 first-time entering students in the 2003–2004 (<1%), 2004–2005 (36%), 2005–2006 (61%), and 2006–2007 (3%) academic years who took the ACT test within three years prior to enrolling in college. FYGPAs were provided by the institutions. HSGPAs were based on students' self-report of grades from up to 23 high school courses in English, mathematics, social studies, and science; students provided the information at the time they registered for the ACT. At the same time, students also provided their race/ethnicity, gender, and annual family income. For race/ethnicity, White, African American, and Hispanic students were investigated. For annual family income, students were classified into the following categories: less than $36,000, $36,000 to $60,000, or greater than $60,000.

Most of the 259 institutions in the sample were four-year public institutions (74%) and had a small percentage of African American and Hispanic students (median percentage of 12% across institutions). A minimum subgroup sample size of 10 was required for inclusion of a postsecondary institution in the analyses. Because it was not possible to construct the true applicant pool for these institutions, an approximate pool was developed. This pool included the enrolled students plus any students from the identified years who sent an ACT score report to at least one of the 259 institutions. For a more detailed description of the sample, see the full report (Sanchez, 2013).

Method. Hierarchical logistic regression models were estimated for predicting attainment of two successive levels of FYGPA: 2.5 or higher and 3.0 or higher. For each of the predictors investigated, alone or in combination, three validity statistics were calculated per institution using the institution-specific total-group optimal cutoff (OC): accuracy rate (AR), success rate (SR), and increase in accuracy rate (ΔAR) to help determine the effectiveness of these measures for making postsecondary admission decisions. (Methodological details can be found in section 11.3.1 on Decision-Based Statistics and in Sawyer, 2010).

For each institution and success level, optimal cutoffs that maximized prediction accuracy for FYGPA were identified for the ACT-C, HSGPA, and joint ACT-C/HSGPA models using a total-group model. The cutoffs were used to simulate the effects of making admission decisions based on ACT-C, HSGPA, or both on student subgroups. Postsecondary institutions do not utilize strict score cutoff values like those used in the present study. The use of strict cutoffs in the present study is a mathematical idealization intended to provide guidance to postsecondary institutions as they decide how best to make admission decisions.

It can be shown that optimal cutoffs also correspond to a 0.50 probability of success for a given model. For the ACT-C and HSGPA joint model, multiple combinations of ACT-C and HSGPA cutoffs corresponding to a probability of success of 0.50 can be identified. Probability distributions that cross 0.50 will yield accuracy rate distributions that increase to a maximum and then decrease. If the probability distribution for an institution does not cross 0.50, the maximum accuracy rate and optimal cutoff indicate that the selection criteria are not useful, and the model is therefore considered a “nonviable” model for an institution. Models for institutions with probability curves crossing 0.50 are referred to here as “viable” models.

For each model investigated, the number of institutions producing viable models varied. The results presented are limited to institutions that produced viable models for the three models examined (i.e., ACT-C, HSGPA, and joint ACT-C and HSGPA models). In the 2.5 or higher and 3.0 or higher success models, 253 and 247 institutions (out of a possible 259 institutions), respectively, produced viable models.

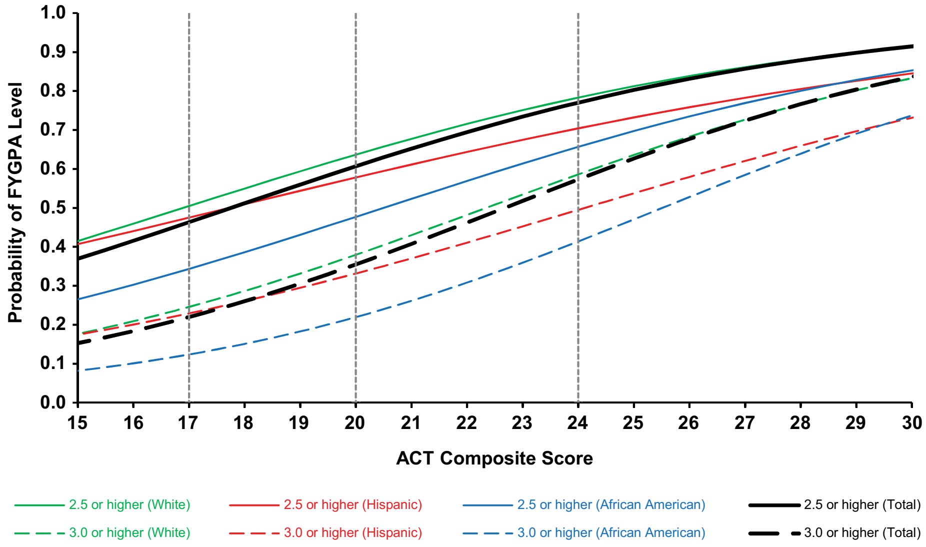
Total-group and subgroup validity statistics were based on the institution’s own frequency distribution of predictor variables and summarized across institutions using median values. Results for each model were based on using the institution-specific total-group cutoffs and applying the cutoff to the subgroup-specific probability and frequency distribution for each institution. These values were used to compare subgroups to examine the differential usefulness in making admission decisions. Typical values of the validity statistics at the total-group optimal cutoffs were compared across student subgroups.

Results. Results for the analyses by race/ethnicity, gender, and income follow.

Race/Ethnicity. For White, African American, and Hispanic students, as ACT-C or HSGPA increased, the probability of success also increased (Figures 11.14 and 11.15). For the two FYGPA levels, White students had higher estimated probabilities of success than African American and Hispanic students had over most of the ACT-C score and HSGPA scales, and Hispanic students tended to have higher estimated chances of success than African American students had. Where differences in over- and underprediction of success existed, they tended to be of greater magnitude when HSGPA was used as the academic predictor then when ACT-C score was used (see Figure 11.15). 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.

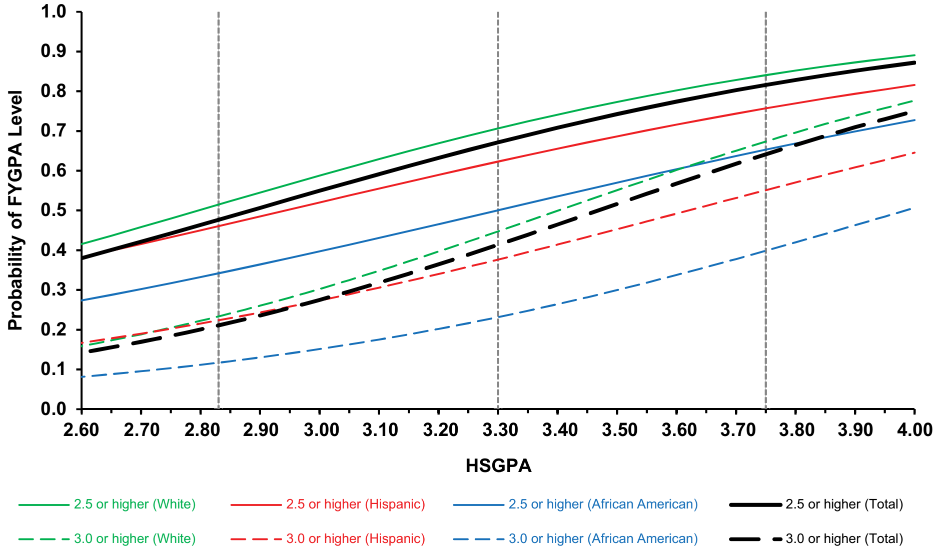
The median probabilities of success across institutions based on a total-group cutoff for racial/ethnic groups tended to show a pattern of underprediction for White students and overprediction for both Hispanic and African American students (see Table 11.21). Across institutions, for the 2.5 or higher success level, Hispanic students showed the least amount of overprediction. African American students, however, showed evidence of moderate overprediction. For the 3.0 or higher success level, the overprediction observed for minority groups increased in magnitude, especially for African American students.

The joint ACT-C and HSGPA model tended to produce the most favorable ARs and SRs, on average, across the racial/ethnic groups (Table 11.21). For the 2.5 or higher FYGPA success level, median ARs were somewhat comparable across racial/ethnic groups. In comparison, for the 3.0 or higher FYGPA success level, median ARs were highest for African American students and lowest for White students. Moreover, for both FYGPA success levels, the increase in accuracy rates (ΔARs) associated with using ACT-C and HSGPA jointly as predictors was greater for African American and Hispanic students than for White students.



Note. The three vertical reference lines represent the first, second, and third quartiles.

Figure 11.14 Estimated probabilities of achieving specific FYGPA levels based on ACT-C score, by race/ethnicity.



Note. The three vertical reference lines represent the first, second, and third quartiles.

Figure 11.15 Estimated probabilities of achieving specific FYGPA levels based on HSGPA, by race/ethnicity.

Table 11.21 Median Statistics for Predicting Specific Levels of FYGPA by Race/Ethnicity across Institutions

Predictor variable	N	Total-group cutoff	Race/ ethnicity	Subgroup-specific probability of success Median (Min/Max)	Maximum accuracy rate (AR) Median (Min/Max)	Increase in AR (ΔAR) Median (Min/Max)	Success rate (SR) Median (Min/Max)	Observed percentage below OC (PB) Median (Min/Max)
2.5 or higher FYGPA								
ACT-C		18	White	0.56 (0.29/0.77)	69 (52/97)	5 (-13/50)	72 (52/97)	29 (0/97)
			African American	0.39 (0.19/0.61)	70 (46/93)	37 (0/86)	52 (18/86)	70 (0/100)
			Hispanic	0.51 (0.2/0.71)	65 (54/86)	21 (-13/72)	59 (26/84)	70 (0/100)
HSGPA	242	2.8	White	0.53 (0.07/0.77)	72 (55/96)	6 (-4/50)	74 (51/96)	25 (0/91)
			African American	0.35 (0.07/0.61)	67 (33/90)	29 (-1/81)	51 (11/82)	55 (0/100)
			Hispanic	0.47 (0.23/0.7)	67 (42/84)	19 (-7/69)	62 (18/82)	55 (0/100)
ACT-C & HSGPA			White	0.52 (0.1/0.73)	73 (59/97)	10 (-4/57)	75 (52/97)	31 (0/92)
			African American	0.37 (0.11/0.85)	73 (45/94)	42 (0/87)	55 (10/86)	70 (0/100)
			Hispanic	0.48 (0.2/0.73)	70 (55/87)	31 (-11/74)	62 (19/83)	70 (0/100)

Note. Multiple combinations of ACT-C score and HSGPA correspond to a 0.50 probability of success for the joint models

Table 11.21 Median Statistics for Predicting Specific Levels of FYGPA by Race/Ethnicity across Institutions—continued

				Subgroup-specific probability of success	Maximum accuracy rate (AR)	Increase in AR (ΔAR)	Success rate (SR)	Observed percentage below OC (PB)
Predictor variable	N	Total-group cutoff	Race/ ethnicity	Median (Min/Max)	Median (Min/Max)	Median (Min/Max)	Median (Min/Max)	Median (Min/Max)
3.0 or higher FYGPA								
ACT-C		23	White	0.54 (0.37/0.75)	71 (62/90)	25 (-2/63)	68 (53/90)	66 (1/99)
			African American	0.36 (0.21/0.74)	86 (57/97)	71 (7/93)	46 (7/75)	93 (14/100)
			Hispanic	0.45 (0.32/0.6)	78 (60/91)	56 (2/83)	53 (16/77)	93 (14/100)
HSGPA	236	3.4	White	0.52 (0.23/0.79)	72 (55/87)	22 (0/60)	68 (51/89)	55 (1/98)
			African American	0.27 (0.15/0.51)	81 (43/98)	64 (0/97)	37 (2/66)	85 (0/100)
			Hispanic	0.42 (0.21/0.59)	75 (53/96)	49 (0/92)	52 (4/77)	85 (0/100)
ACT-C & HSGPA			White	0.51 (0.36/0.69)	75 (57/90)	30 (1/69)	70 (54/90)	62 (2/97)
			African American	0.32 (0.02/0.61)	87 (61/100)	73 (14/99)	48 (1/87)	92 (19/100)
			Hispanic	0.43 (0.04/0.6)	81 (63/98)	61 (6/96)	55 (3/80)	93 (18/100)

Note. Multiple combinations of ACT-C score and HSGPA correspond to a 0.50 probability of success for the joint models

Gender. For both males and females, as ACT-C or HSGPA increased, the estimated probability of attaining the two FYGPA success levels also increased (figures provided on pp. 26–27 of Sanchez, 2013). Moreover, regardless of the level of success examined, females had a higher probability of success than males. There also appeared to be a trend of greater overprediction for males than underprediction for females. As shown in Table 11.22, using a total-group cutoff score underpredicted the probability of success for females and overpredicted the probability of success for males for both success levels. Across institutions, the use of ACT-C alone resulted in slightly larger differential prediction than when HSGPA was used in isolation.

Table 11.22 Median Statistics for Predicting Specific Levels of FYGPA by Gender across Institutions

Predictor variable	N	Total-group cutoff	Gender	Subgroup-specific probability of success Median (Min/Max)	Maximum accuracy rate (AR) Median (Min/Max)	Increase in AR (ΔAR) Median (Min/Max)	Success rate (SR) Median (Min/Max)	Observed percentage below OC (PB) Median (Min/Max)
2.5 or higher FYGPA								
ACT-C		18	Female	0.56 (0.40/0.68)	73 (59/97)	8 (0/64)	75 (58/97)	33 (0/96)
			Male	0.45 (0.33/0.61)	69 (55/92)	16 (0/73)	62 (33/92)	40 (0/100)
HSGPA	253	2.8	Female	0.53 (0.16/0.61)	73 (57/96)	6 (0/51)	75 (53/96)	24 (0/89)
			Male	0.47 (0.18/.059)	70 (56/91)	13 (-1/64)	66 (48/91)	35 (0/94)
ACT-C & HSGPA			Female	0.52 (0.07/0.61)	75 (60/97)	12 (0/65)	76 (53/97)	33 (0/93)
			Male	0.45 (0.10/0.57)	72 (57/92)	20 (0/74)	66 (45/92)	44 (0/99)
3.0 or higher FYGPA								
ACT-C		23	Female	0.59 (0.36/0.73)	74 (63/92)	27 (0/76)	74 (56/93)	68 (0/99)
			Male	0.43 (0.32/0.57)	74 (59/94)	43 (0/89)	58 (20/91)	74 (0/100)
HSGPA	247	3.4	Female	0.52 (0.32/0.61)	73 (62/93)	24 (0/62)	68 (47/93)	54 (1/98)
			Male	0.46 (0.32/0.54)	74 (59/92)	38 (0/78)	60 (36/94)	66 (1/99)
ACT-C & HSGPA			Female	0.53 (0.28/0.6)	77 (66/92)	32 (0/78)	73 (53/93)	62 (0/99)
			Male	0.44 (0.3/0.53)	78 (64/95)	45 (0/90)	62 (32/93)	73 (1/100)

Note. Multiple combinations of ACT-C score and HSGPA correspond to a 0.50 probability of success for the joint models.

For the FYGPA 2.5 or higher success level, using a total-group cutoff resulted in higher median ARs and SRs for females than for males, regardless of the predictor combination used. At the 3.0 or higher level, while the median SRs were higher for females than for males, median ARs were more similar between males and females. For both success levels, typical ΔARs were considerably larger for males than for females, and a smaller percentage of males were at or above the total-group cutoff than were females. For both success levels the joint ACT-C and HSGPA model tended to produce more favorable ARs and SRs, on average, for both males and females.

Income. For lower-, middle-, and higher-income students, as ACT-C or HSGPA increased, the estimated probability of achieving the two FYGPA levels also increased. For both success levels, when either ACT-C or HSGPA was used as the sole academic predictor, the estimated probabilities of success for lower-income students tended to be lower than the estimated probabilities for middle-income students, and both tended to be lower than the estimated probabilities of higher-income students (figures provided on p. 31 of Sanchez, 2013). The median probability of success at the total-group cutoff for lower- and higher-income students tended to be over- and underpredicted, respectively (see Table 11.23). Relatively little evidence of over- or underprediction was observed for middle-income students.

For the 2.5 or higher FYGPA success level, as income level increased, typical ARs also increased slightly, with this finding being more pronounced for the HSGPA alone model. For the 3.0 or higher level, as income increased, typical ARs tended to decrease somewhat. In comparison, typical ΔARs were considerably larger for lower-income students than for higher-income students at both FYGPA success levels. The joint ACT-C and HSGPA model tended to produce slightly more favorable ARs and SRs, on average, across the income groups for both success levels.

Table 11.23 Median Statistics for Predicting Specific Levels of FYGPA by Income across Institutions

Predictor variable	N	Total-group cutoff	Income	Subgroup-specific probability of success Median (Min/Max)	Maximum accuracy rate (AR) Median (Min/Max)	Increase in AR (ΔAR) Median (Min/Max)	Success rate (SR) Median (Min/Max)	Observed percentage below OC (PB) Median (Min/Max)
2.5 or higher FYGPA								
ACT-C		18	Lower	0.49 (0.44/0.53)	69 (53/90)	18 (0/81)	63 (48/90)	50 (0/99)
			Middle	0.52 (0.48/0.57)	70 (55/95)	10 (0/76)	70 (54/95)	37 (0/98)
			Higher	0.55 (0.5/0.61)	71 (55/97)	5 (0/67)	73 (57/97)	27 (0/98)
HSGPA	253	2.8	Lower	0.47 (0.39/0.54)	68 (51/90)	12 (-1/60)	63 (45/90)	36 (0/95)
			Middle	0.49 (0.34/0.55)	72 (55/95)	10 (0/52)	72 (54/95)	29 (0/90)
			Higher	0.53 (0.37/0.62)	74 (57/96)	6 (0/41)	77 (56/96)	24 (0/90)
ACT-C & HSGPA			Lower	0.47 (0.09/0.59)	72 (51/90)	22 (0/79)	65 (49/91)	49 (0/99)
			Middle	0.50 (0.1/0.54)	74 (55/96)	14 (0/75)	73 (54/96)	36 (0/99)
			Higher	0.53 (0.08/0.59)	75 (56/97)	10 (0/66)	77 (55/97)	30 (0/98)

Note. Multiple combinations of ACT-C score and HSGPA correspond to a 0.50 probability of success for the joint models.

Table 11.23 Median Statistics for Predicting Specific Levels of FYGPA by Income across Institutions—continued

Predictor variable	N	Total-group cutoff	Income	Subgroup-specific probability of success Median (Min/Max)	Maximum accuracy rate (AR) Median (Min/Max)	Increase in AR (ΔAR) Median (Min/Max)	Success rate (SR) Median (Min/Max)	Observed percentage below OC (PB) Median (Min/Max)
3.0 or higher FYGPA								
ACT-C		23	Lower	0.48 (0.43/0.53)	76 (60/92)	46 (0/83)	61 (41/86)	81 (0/100)
			Middle	0.52 (0.48/0.6)	74 (63/92)	33 (0/77)	67 (52/93)	71 (0/99)
			Higher	0.54 (0.5/0.65)	72 (61/96)	24 (0/68)	69 (54/96)	63 (0/97)
HSGPA	247	3.4	Lower	0.43 (0.35/0.59)	72 (53/89)	38 (0/70)	54 (39/90)	68 (2/99)
			Middle	0.49 (0.46/0.58)	74 (58/94)	29 (0/66)	65 (52/95)	58 (2/98)
			Higher	0.54 (0.47/0.64)	73 (58/97)	22 (0/61)	70 (56/97)	54 (1/97)
ACT-C & HSGPA			Lower	0.45 (0.37/0.56)	77 (60/92)	47 (0/84)	61 (44/90)	76 (0/99)
			Middle	0.50 (0.44/0.55)	77 (62/93)	36 (0/78)	69 (51/95)	66 (0/99)
			Higher	0.53 (0.38/0.69)	76 (59/95)	10 (0/69)	72 (59/97)	60 (0/96)

Note. Multiple combinations of ACT-C score and HSGPA correspond to a 0.50 probability of success for the joint models.

Summary. Across student subgroups, the joint use of ACT-C and HSGPA resulted in greater prediction accuracy than when either predictor was used alone. Furthermore, the use of a total-group cutoff score for both ACT-C and HSGPA slightly overpredicts the probability of success of Hispanic and African-American students, males, and lower-income students. Both ACT-C and HSGPA slightly underpredict the probability of success of White students, females, and higher-income students. These findings suggest, therefore, that African American, Hispanic, and lower-income students 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. For further details on both studies, see the full ACT Research Reports (Sanchez, 2013; Radunzel & Noble, 2013).

In conclusion, the use of multiple measures helps to capture a more holistic view of student readiness. 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 FYGPA prediction model that already included ACT Composite score and HSGPA helped to increase the predictive validity and reduce the amount of differential prediction by gender in FYGPA estimates.

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 11.8% (2004–2005), and the percentages have remained above 13% from 2005 to 2011 (Snyder & Dillow, 2013). 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
- Reader's script administration

Average scores for those tested in 2013–2014 are shown in Table 11.24. On average, students with disabilities testing with accommodations earn lower scores than those of students from the overall ACT-tested population. The few exceptions are students with motor impairments and psychiatric disorders (e.g., mood or anxiety).

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). This section describes a more recent study by Huh and Huang (2016) that was conducted to examine this issue.

Data and method. ACT accommodation records from 433,694 students who were given some type of testing accommodation from January 2009 to December 2013 were collected. First-year college outcome data were provided by postsecondary institutions that participated in various ACT research services or partnerships. After ACT accommodation records were matched to first-year college outcome data, the scores of 1,766 students (enrolled at 143 postsecondary institutions) who tested with accommodations and had a valid FYGPA and a valid ACT Composite score were retained for the analyses. Scores for 187,100 students at these institutions who tested without accommodations were also retained for the study. Only a few disability groups had sufficient samples of students testing with accommodations. Specifically, the analyses included two disability groups (382 students with an attention deficit disorder and 883 students with a reading disability) and two extended-time accommodations groups (652 students with up to triple time on each test over multiple days and 623 students with up to time-and-a-half time on each test over multiple days).

Consistent with Ziomek and Andrews (1996), institution-specific regression equations for the total group were calculated. Institution-specific total-group regression parameters were then applied separately to students testing with and without accommodations to obtain their predicted college GPAs.

Table 11.24 Average ACT Scores for Students Tested with Accommodations in 2013–2014

Reference group	Number of students	Average ACT score				
		English	Mathematics	Reading	Science	Composite
Learning Disability						
Mathematics Disorder	3,585	14.3	15.4	16.4	15.5	15.5
Reading Disorder	31,753	13.7	16.5	16.3	16.9	16.0
Writing Disorder/Written Expression	938	16.7	19.2	18.7	19.4	18.6
Speech/Language Disorder	251	15.6	17.7	17.5	18.3	17.4
Physical/Sensory disabilities						
Hearing Impairment	1,132	13.2	16.7	16.3	17.3	16.0
Motor Impairment[1]	719	21.1	20.5	23.4	21.9	21.8
Visual Impairment[2]	869	19.0	19.2	21.5	19.7	20.0
Other Physical/Sensory Disability	218	18.5	19.0	19.0	19.8	19.2
Psychological Disability						
ADD/ADHD	14,449	18.2	18.8	19.9	19.4	19.2
Psychiatric Disorder[3]	937	23.9	22.4	25.6	23.7	24.0
Emotional/Behavioral Disorder	2,294	15.3	16.5	17.3	16.8	16.6
Autism Spectrum Asperger's Disorder	1,314	18.6	18.8	19.6	19.8	19.4
Traumatic Brain Injury	81	18.1	19.1	19.3	19.5	19.0
Other[4]	8,779	12.2	15.2	14.6	15.2	14.4
All ACT-tested graduates, 2014	1,845,787	20.3	20.9	21.3	20.8	21.0

Notes. [1]– e.g., Cerebral Palsy, Muscular Dystrophy. [2]– e.g., 20/100 corrected Visual Acuity. [3]– e.g., Mood or Anxiety. [4]– Including Mental or Intellectual Disability.

Source: Ndum, Radunzel, & Westrick (2016)

Results. When jointly using ACT Composite scores and HSGPAs to predict FYGPAs, the mean error of prediction (e.g., observed FYGPA minus predicted FYGPA) for the regular-tested group of students who tested without accommodations was 0.00. The predicted FYGPAs of students testing with accommodations tended to be slightly higher (0.05), on average, than their actual FYGPAs. Residuals for the predicted FYGPAs were larger when using either ACT Composite scores or HSGPAs alone. The correlation between predicted FYGPA and actual FYGPA for all special-tested students was .45, as compared to .56 for regular-tested students.

Summary. 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 scores 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.

11.4 Making Course Placement Decisions

The ACT tests were expressly designed to facilitate placement in 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 on 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.

11.4.1 Placement Validity Argument Based on ACT Content

A validity argument for a placement test can, in part, be based on subject matter content. The ACT test battery 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 contents of the ACT are 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 the ACT test(s) 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, it is unlikely that ACT scores will 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 course work and grades, scores on locally developed placement tests, or noncognitive measures, in addition to ACT scores in 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.

11.4.2 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. This information provides validity evidence for using ACT scores for placement.

Data and method. Grade data were from entry-level courses at two-year and four-year institutions and included several different course types. The institutions participated in the ACT Course Placement Service, ACT Prediction Service, or in special studies (e.g., statewide placement studies) prior to 2014. The results of these analyses were summarized across institutions by course type.

Within each institution, courses that had at least 50 students who had completed the corresponding ACT test and had earned a course grade were included in the analysis. The sample for each course was weighted to match the population of ACT-tested enrollees at each institution on gender, race/ethnicity, ACT Composite score level, and HSGPA level. ACT-tested enrollees from the entering freshmen classes of 2013–2015 were identified using enrollment records from the National Student Clearinghouse and the ACT Class Profile Service.

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 analysis. At each ACT score, the success and accuracy rates were estimated from the probabilities of success obtained from the logistic regression model (see section 11.3.1 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 predictions of course success. When examined across a range of possible cutoff scores for a given institution, the accuracy rate will typically peak at a specific score and then decrease as the score increases further.

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 large 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, 2012b; 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. Table 11.25 provides the summarized results for 17 courses. For all courses, the median accuracy rate at the optimal cutoff score was at least 62%. Thus, a typical institution using the ACT optimal cutoff score from their data could expect that 62% or more of the placement decisions that are made would be correct decisions. 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%. Thus, if all students were allowed into the course, the expected accuracy rate would be 53%.

Mathematics, social science, and natural science courses tended to show higher increases in accuracy rates than English courses. 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 after using ACT scores for placement into English courses. Results from other ACT research suggest this phenomenon occurs regardless of the placement variable (e.g., standardized tests, high school grades, locally developed placement tests, or performance assessments).

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 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.

Table 11.25 Decision-Based Validity Statistics for Course Placement Using ACT Scores (Success criterion = B or higher grade)

Course type	ACT score	Number of institutions	Median cut score*	Median logistic slope	Maximum accuracy rate			Increase in accuracy rate			Success rate		
					Q_1	Med.	Q_3	Q_1	Med.	Q_3	Q_1	Med.	Q_3
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
Mathematics courses													
Elementary Algebra	Mathematics	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.

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 11.4.5.

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.

11.4.3 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). This section describes a study that 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).

Data and method. Using course grade data from 17 cohort years (1996–2012) representing more than 500,000 student outcomes at more than 200 two-year institutions, hierarchical logistic regression models were developed to estimate the conditional probabilities of success in a course as a function of the corresponding ACT multiple-choice test scores and HSGPAs and their interaction, accounting for institution attended. Models for five courses were estimated. Institutions reported the courses as either standard (credit earned) or developmental/remedial (no credit earned). In standard courses (Composition I and College Algebra), success was defined as earning a course grade of B or higher. In the developmental courses (Reading, Elementary Algebra, and Intermediate Algebra), success was defined as earning a grade of C or higher because these courses often use pass/fail grading.

Results. Figures 11.16 and 11.17 illustrate the value of using multiple measures when estimating a student's likelihood of course success. Figure 11.16 plots the probability of earning a grade of B or higher in English Composition I at two-year institutions given a student's ACT English score and HSGPA. At each ACT English score point, the probability of success varies depending on HSGPA. If only ACT scores were available, there would be only one probability curve, and students with the same score would have the same estimated probability of success. Similarly, if only HSGPA were used to predict success, students with the same HSGPA would have the same estimated probability of success. For example, a student with an ACT English score of 15 and a HSGPA of 3.0 has a .46 probability of earning a grade of B or higher at a typical institution. However, if the student had an ACT English score of 20 and a HSGPA of 3.0, the probability would be .53, and if the student had an ACT English score of 20 and a HSGPA of 3.5, the probability would be .67. These results demonstrate how a high HSGPA can “compensate” for a low ACT score, and vice versa. Similar patterns can be seen in Figure 11.17, which displays probability curves for earning a grade of B or higher in College Algebra courses given a student’s ACT mathematics score and HSGPA. As demonstrated by these figures, institutions can more accurately predict a student’s chance of success in college when they use more than one measure. Refer to the full report for additional information (Westrick, 2016).

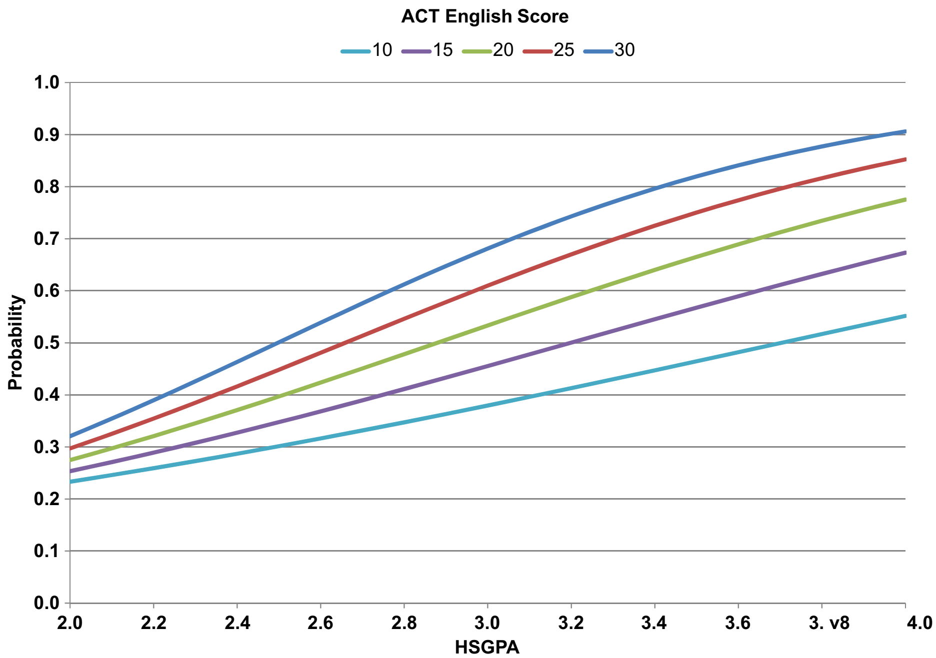


Figure 11.16 Probability of earning a grade of B or higher in English Composition I at two-year institutions, given ACT English score and HSGPA.

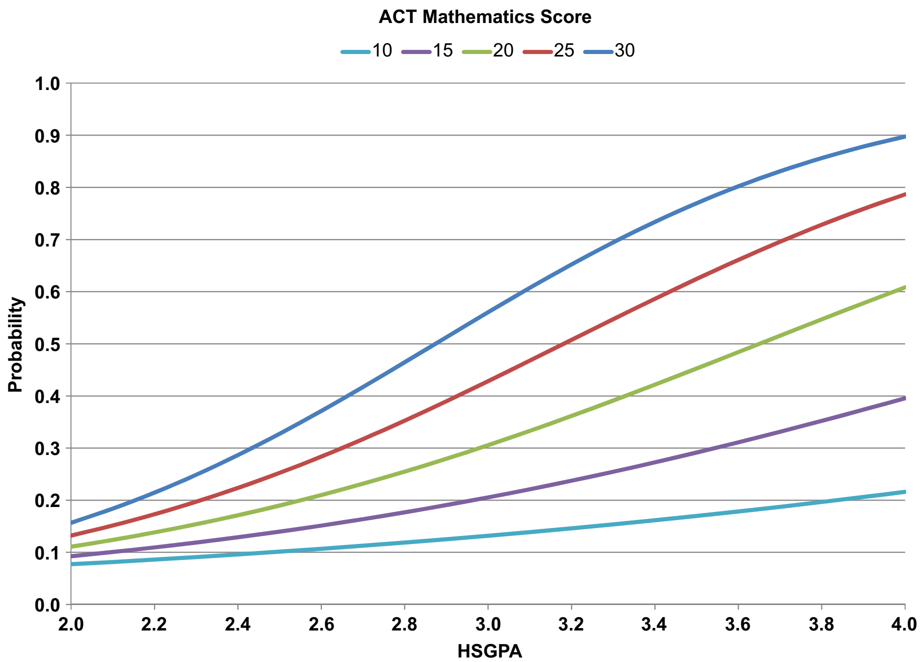


Figure 11.17 Probability of earning a grade of B or higher in College Algebra at two-year institutions, given ACT mathematics score and HSGPA.

Supplemental analyses using the same data set were conducted to obtain the median accuracy rates, the median increase in accuracy rates, and observed success rates for English Composition I and College Algebra. Results are presented in Table 11.26. The accuracy rates indicate the estimated percentage of correct placement decisions based on using the predictor variable(s). The increase in accuracy rates indicate the increment in the percentage of correct placement decisions when using the predictor(s) for placement compared to not using any predictor variables for course placement (i.e., all students were placed into the standard course). In 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. 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.

Summary. The use of either ACT scores or HSGPA for placement purposes results in accuracy rates higher than the expected accuracy rates if all students were allowed to enroll in the standard course. However, the joint use of ACT scores and HSGPA results in higher accuracy rates.

Table 11.26 Median Placement Statistics for ACT Scores and HSGPA as Predictors at Community Colleges

Course type	Number of institutions	Number of students	Predictor variable	Median Accuracy Rate	Median Increase in Accuracy Rate	Observed Success Rate
English Composition	259	288,266	ACT English	63.3	4.9	60.6
		256,110	HSGPA	66.7	8.3	61.2
		256,110	ACT English, HSGPA, & ACT English × HSGPA	66.8	7.9	61.2
College Algebra	182	132,850	ACT Math	66.5	25.9	42.2
		119,228	HSGPA	67.7	19.9	43.2
		119,228	ACT Math, HSGPA, & ACT Math × HSGPA	68.6	24.5	43.2

Notes. 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.

11.4.4 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 language learners, students with disabilities, racial/ethnic minority students, and low-income students. More 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.

Data and method. The data used in this study were the same as those used to update the ACT College Readiness Benchmarks (Allen, 2013). Briefly, data came from colleges or groups of colleges that participated in ACT's research services, including the Course Placement Service and Prediction Service. Results were based on 96,583 students from 136 colleges for English Composition I, 70,461 students from 125 colleges for College Algebra, and 41,651 students from 90 colleges for Biology. Six different courses were considered for the social science analyses: American History, Other History, Psychology, Sociology, Political Science, and Economics. Results for the social science courses were based on 130,954 students from 129 colleges.

The information used to identify the demographic groups was provided voluntarily by students via the ACT test registration process. Identification of English language learners was based on whether English was the language most commonly spoken in the student's home; 2% to 3% of the students in the course samples were classified as English language learners. When registering for the ACT, students were asked, "Do you have a disability that requires special provisions from the educational institution?" Positive responses to this question were used to identify students with disabilities. Examinees with documented disabilities may take the ACT with special accommodations. Options include standard testing time with accommodations, 50% extended testing time, and special testing at school that can allow more than 50% extended time. Students' ACT scores obtained from extended testing time were not used in analyses. Therefore, some students with disabilities were excluded from the analysis. For reference, among students in the 2015 ACT-tested graduating class who reported having a disability that requires special testing provisions, about 25% only took the ACT with extended time. Four to five percent of students in the course samples were classified as students with disabilities. Racial/ethnic minorities included African American, Native American, Hispanic, Native Hawaiian, students of multiple races, and students of other races (not including White and Asian); 20% to 24% of the students in the course samples were classified as racial/ethnic minority. The 24% to 28% of students reporting an annual family income of $36,000 or lower were classified as low-income.

Success in a course was defined as earning a grade of B or higher. Hierarchical logistic regression was used to model within each college the probability of success in a course as a function of ACT test score. The 50% cut scores for the demographic groups were derived from the fixed effect parameter estimates from the regression models.

Results. For all demographic groups and subject areas, there was a positive relationship between ACT score and probability of success in the college course (see Figure 11.18 for College Algebra). The slope for students with disabilities was consistently flatter than those for most other groups and the total group of students (see Table 2 from Allen, 2016b). The slope for English language learners was also flatter than those for the total group in all subject areas. Slopes for racial/ethnic minority and low-income students were more similar to those obtained for the total group.

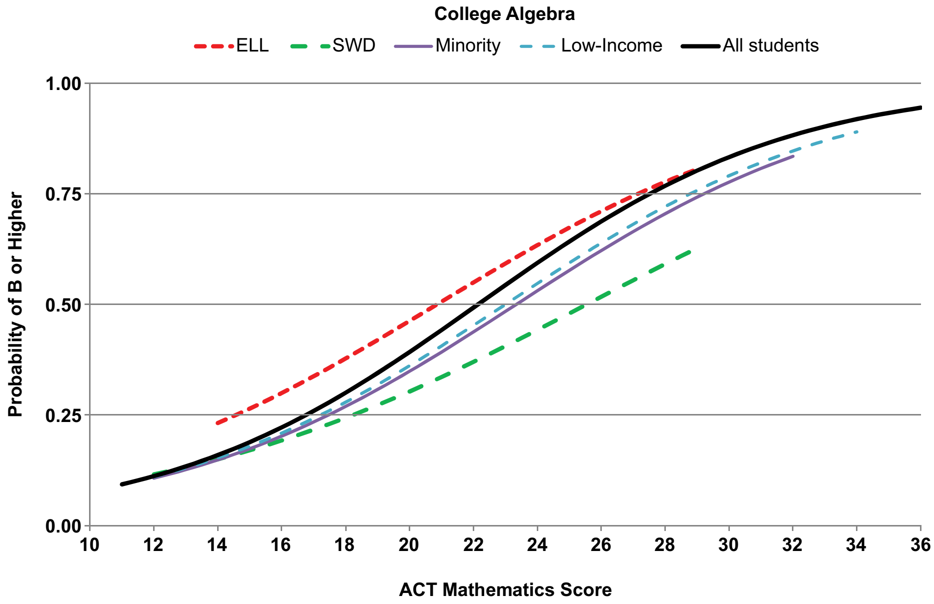


Figure 11.18 Probability of earning a grade of B or higher in College Algebra by ACT mathematics score and student demographic group (ELL is for English language learners; SWD is for students with disabilities).

Table 11.27 provides estimates of the 50% success cut scores for each group and subject area as compared to the ACT College Readiness Benchmarks of 18, 22, 22, and 23 in English, mathematics, reading, and science, respectively. Note that the cut scores for the demographic groups represent the typical cut score across institutions, but they do not incorporate the additional steps used to derive the ACT College Readiness Benchmarks (e.g., weighting the sample to be nationally representative). Across subject areas, the ACT scores required to have at least a 50% chance of success were lower for English language learners and higher for students with disabilities, racial/ethnic minorities, and low-income students as compared to the total group. When the 50% success cut score for a group is higher than the 50% cut score for the total group, overprediction occurs for that group. That is, at the 50% cut score for the total group, the chance of success is lower than 50% for the demographic group of interest. Similarly, underprediction occurs when the 50% success cut score for a demographic group is lower than the 50% cut score for the total group.

Table 11.27 Scores Associated with at Least a 0.50 Probability of Success for Student Groups Used to Develop the ACT College Readiness Benchmarks

Student group	ACT score (college course)			
	English (English Composition)	Mathematics (College Algebra)	Reading (Social Science)	Science (Biology)
English language learners	16	21	21	23
Students with disabilities	21	26	25	26
Racial/ethnic minority	19	23	25	25
Low-income	18	23	24	24
ACT College Readiness Benchmark/All students	**18**	**22**	**22**	**23**

Summary. The results of this study are consistent with prior research showing slight underprediction for English language learners (Mattern, Patterson, Shaw, Kobrin, & Barbuti, 2008; Patterson & Mattern, 2012) and slight overprediction for students with disabilities (Huh & Huang, 2016; Ziomek & Andrews, 1996), racial/ethnic minority students (Lorah & Ndum, 2013; Noble, Crouse, & Schultz, 1996; Sanchez, 2013; Sawyer 1985), and low-income students (Lorah & Ndum, 2013; Sanchez, 2013) 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 patterns of over-/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 (i.e., 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).

11.4.5 Methods for Setting Cutoff Scores

Institutions have unique placement needs that require locally developed cutoff scores rather than the median optimal cutoff scores shown in this section. 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, as used by the ACT Course Placement Service, 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 11.28 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 .46. That is, 46 out of 100 students with an ACT mathematics score of 18 would be expected to earn a grade of B or higher grade in Mathematics 100. This information is also shown graphically in Figure 11.19.

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 11.28 Probability of Success in Mathematics 100, Given ACT Mathematics Score

ACT mathematics 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

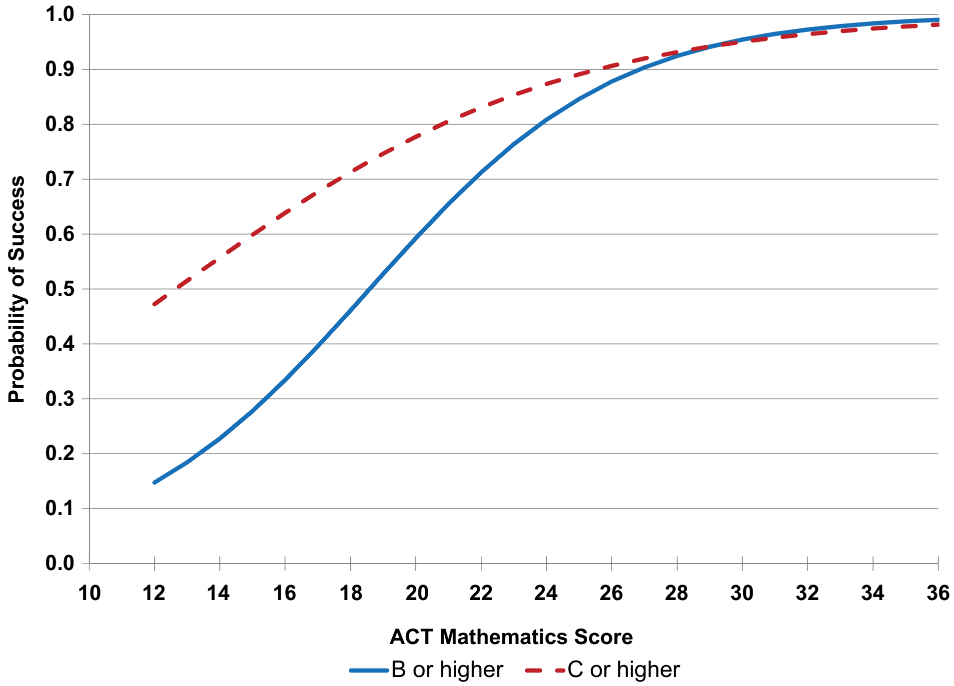


Figure 11.19 Probability of success in Mathematics 100, given ACT mathematics score.

Table 11.29 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 .50. In Tables 11.28 and 11.29, 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 11.29, 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 11.29 Decision-Based Statistics for Placement Based on ACT Mathematics Score

		B or higher		C or higher	
ACT mathematics score	Percent placed in lower-level course	Estimated accuracy rate (in percent)	Estimated success rate (in percent)	Estimated accuracy rate (in percent)	Estimated success rate (in percent)
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) should be related to college and/or course outcomes.

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 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 by using the scores from the ACT national norms or Table 11.25. This is particularly helpful when local normative data are not available. For example, the normative data provided in Appendix A might be used to set local cutoff scores based on the scores earned by a nationally representative sample of ACT–tested students. 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. The Course Placement Service provides a convenient way for institutions to validate and determine appropriate cutoff scores.

11.4.6 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. By participating in the ACT Course Placement Service, institutions receive reports including tables that illustrate the effectiveness of score cutoffs for course placement. These reports can help institutions determine initial score cutoffs, and then reports on new samples of students can be used to evaluate these cutoffs in subsequent years.

11.5 Evaluating Students' Likelihood of College Success

Sections 11.3 and 11.4 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 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.

11.5.1 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 ACT Benchmarks met (see Chapter 8 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 (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. First-year course grades data spanned multiple years from various postsecondary institutions who participated in ACT's Course Placement Service. 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 ACT 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 11.20; by 23 to 29 percentage points); (b) earn a B or higher grade in first-year college courses (Figure 11.21; by 18 to 27 percentage points); (c) achieve a FYGPA of 3.0 or higher (Figure 11.22; by 23 to 27 percentage points), and (d) remain enrolled at the same institution in year two (Figure 11.23; by 6 to 9 percentage points). Moreover, as the number of ACT Benchmarks increased, students' likelihood of success also increased for each of the first-year outcomes examined (Table 11.30). 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.

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 ACT 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.

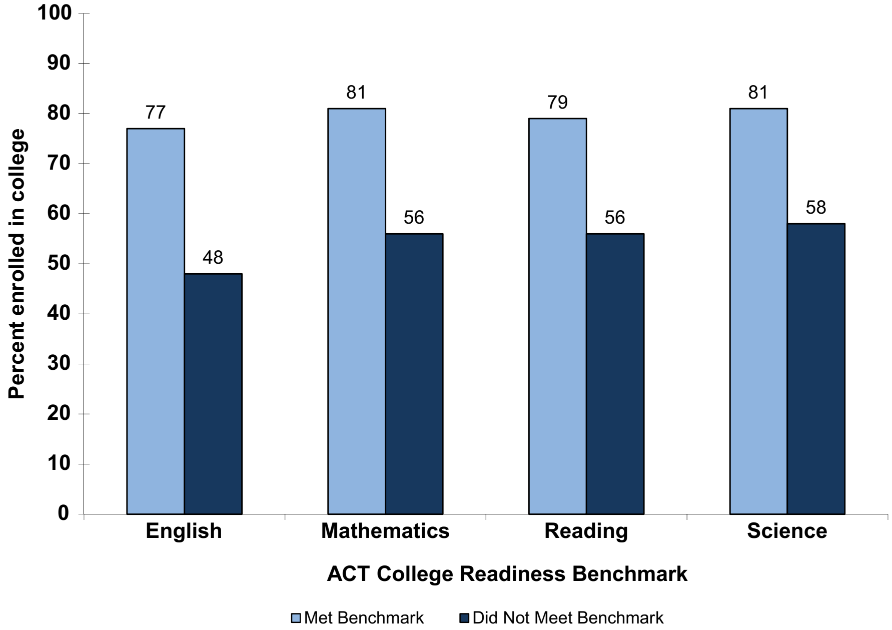


Figure 11.20 College enrollment rates by ACT College Readiness Benchmark attainment.

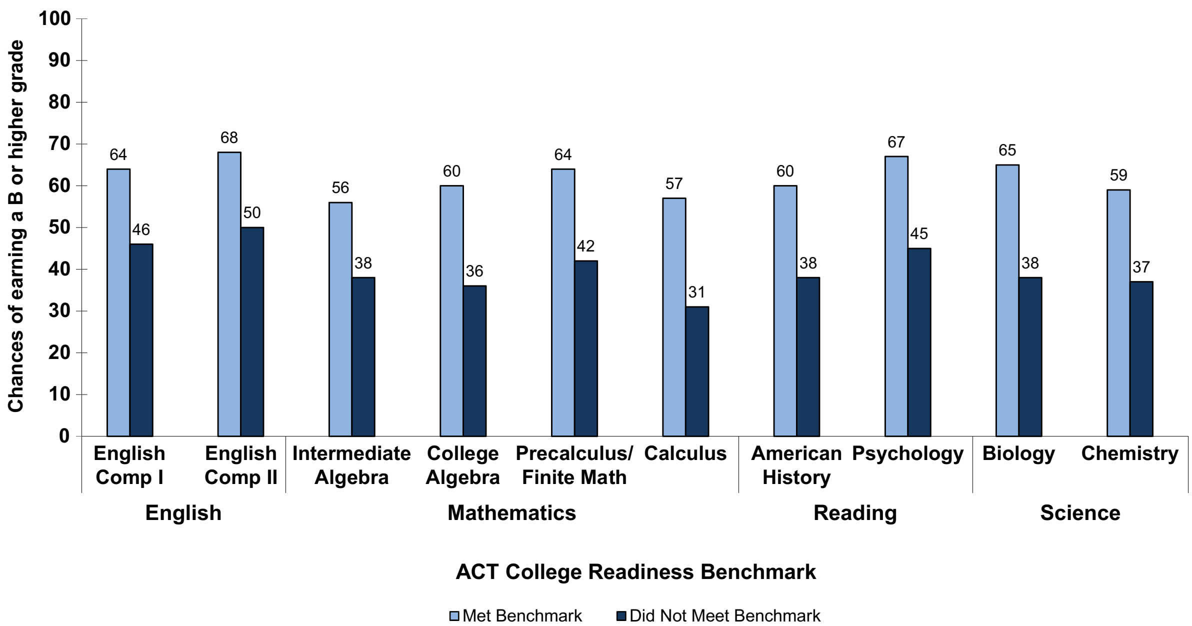


Figure 11.21 Students' chances of earning a B or higher grade in first-year college courses by ACT College Readiness Benchmark attainment at a typical institution.

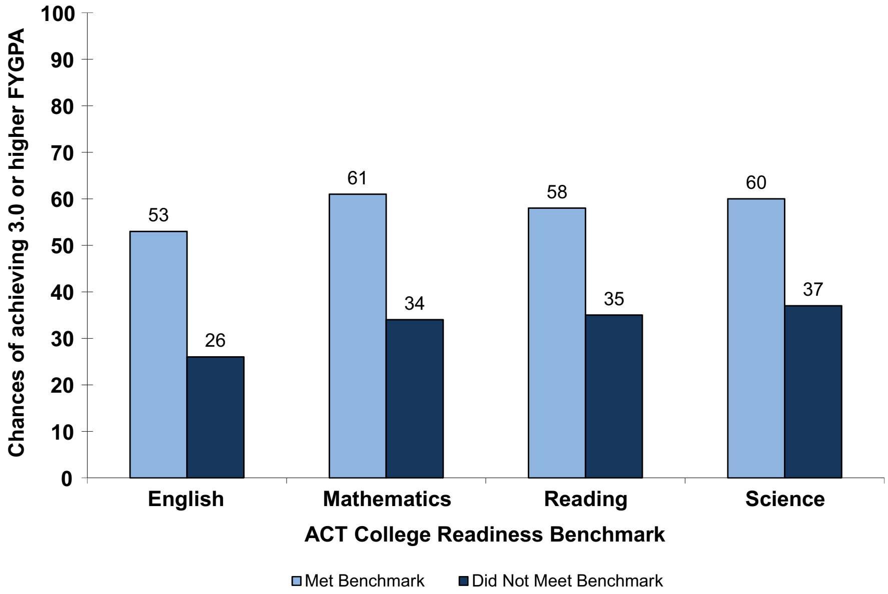


Figure 11.22 Students' chances of achieving a 3.0 or higher FYGPA by ACT College Readiness Benchmark attainment at a typical institution.

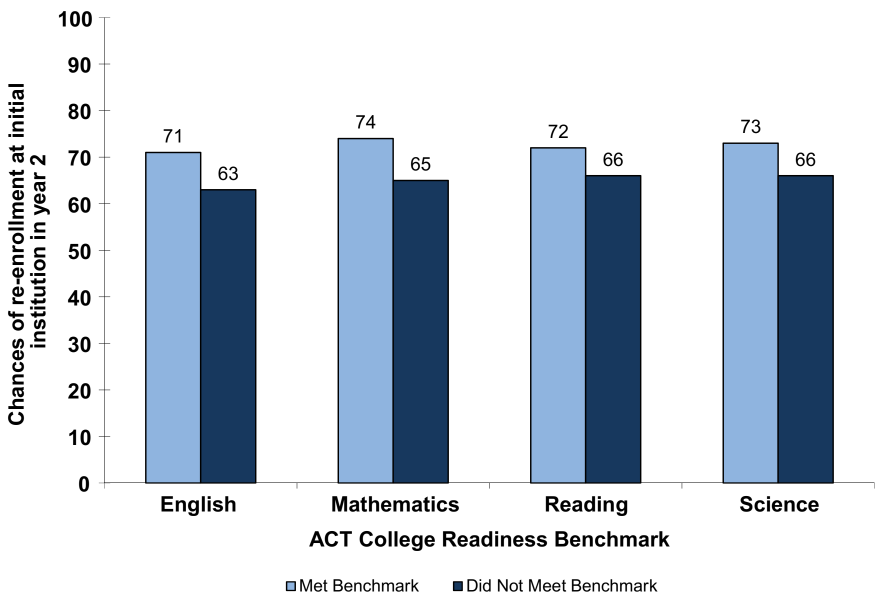


Figure 11.23 Students' chances of remaining enrolled at the initial institution in year two by ACT College Readiness Benchmark attainment.

Table 11.30 First-Year College Outcomes by Number of ACT College Readiness Benchmarks Met

	Number of ACT Benchmarks met				
Outcome	0	1	2	3	4
Enrollment	45	66	73	78	83
B or higher grade in course					
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

11.5.2 Statistical Relationships between ACT and ACT CAAP Scores

The previous section showed that students who are better prepared academically, as measured by meeting the ACT Benchmarks, are more likely to succeed during their first year of college than are underprepared 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 CAAP scores and ACT scores/ACT Benchmark attainment were examined for second-year college students.

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. ACT CAAP is a standardized assessment program that enables postsecondary institutions to assess, evaluate, and enhance the outcomes of their general education programs. ACT CAAP offers six independent test modules: reading, science, critical thinking, mathematics, writing skills, and writing essay (ACT, 2015b). The ACT CAAP assessment was taken by students between the academic years 2008–2009 and 2014–2015. Because of the modular nature of ACT CAAP, not all students with ACT/CAAP matched records 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 CAAP testing were also available as an indicator of college achievement. College readiness was measured by ACT College Readiness Benchmark attainment (see Chapter 8 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 ACT Benchmark attainment relate to ACT CAAP scores and cumulative college GPA in the second year.

Results. ACT scores were strongly correlated with ACT CAAP performance (Table 11.31). In addition, students meeting the ACT College Readiness Benchmarks in high school had higher average ACT CAAP scores than students not meeting the ACT Benchmarks (Table 11.32). This pattern was observed in all four content areas. The difference in average ACT CAAP scores was as much as 6.6 points. Moreover, as shown in Figure 11.24, students who met the ACT College Readiness 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 11.31 ACT/CAAP Test Score Correlations

ACT/CAAP content area			
English/Writing Skills	Mathematics/Mathematics	Reading/Reading	Science/Science
.77	.73	.70	.67

Table 11.32 Average ACT CAAP Test Score by ACT Benchmark Attainment

	ACT Benchmark Attainment	
ACT/CAAP content area	Met	Not met
English/Writing Skills		
Mean (SD)	65.0 (4.0)	58.4 (3.5)
Number of students	8,418	2,803
Mathematics/Mathematics		
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

Note. SD = standard deviation.

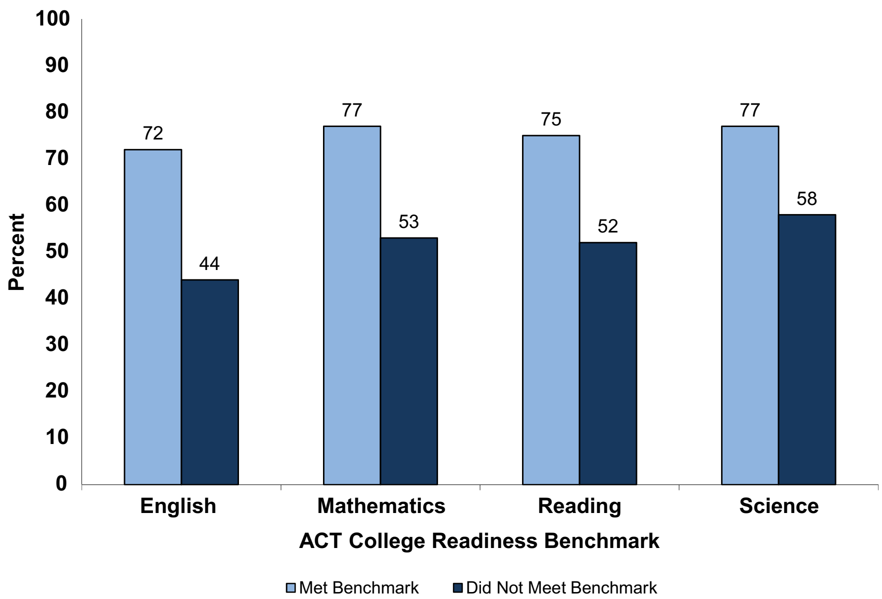


Figure 11.24 Percentages earning a cumulative college GPA greater than 3.00 by ACT College Readiness Benchmark attainment for students taking ACT CAAP during sophomore year and the ACT in high school.

11.5.3 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. The results of this study are summarized in this section.

Data and method. Enrollment information, including enrollment patterns, grades, and majors, were obtained from 87 colleges and universities from four states. All colleges were bachelor's-level-degree-granting institutions. Some colleges provided only one semester of data, while others provided several years of college data. The data included first-time freshmen enrolled between 1994 and 2003; only students with valid ACT scores who had completed the ACT Interest Inventory were included in the analyses. The resulting sample size was 308,500 ACT-tested students who had at least first-year college enrollment data available.

College outcomes included cumulative college GPA at the end of the first academic year, at the end of the second academic year, and at graduation after five academic years. 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. In the models, ACT scores were group-mean centered within institution.

Results. The results of the HLM analyses for college GPA are summarized in Table 11.33. In the table, the fixed-effect columns report the model coefficients estimating the relationship between ACT scores and college GPA, and the random-effect column reports the variance across colleges associated with each variable. For each college GPA outcome, both the fixed effects and random effects were statistically significant ($p < .001$). Mean college GPAs varied significantly across colleges (the intercept), and ACT scores were significantly related to college GPA at various time points (the slope, labeled as ACT in the table). For each model, the proportion of within-college variance (labeled Level-1) accounted for ranged from .11 to .15.

Table 11.33 Summary of Hierarchical Linear Modeling Regression on College GPA

Variable	Fixed effect: Coefficient	Fixed effect: Standard error	Random effect variance	R^2
First-year college GPA (N = 72,648)				
Intercept	273.47*	2.78	341.69*	
ACT	6.55*	0.40	6.26*	
Level-1			5120.49	.11
Second-year college GPA (N = 51,012)				
Intercept	291.44*	2.89	243.38*	
ACT	6.49*	0.35	2.74*	
Level-1			2957.51	.15
Graduation college GPA (N = 15,882)				
Intercept	314.53*	1.49	106.54*	
ACT	5.34*	0.91	0.95*	
Level-1			1884.49	.15

Note. College GPA ranged from 0 to 425; $*p < .001$.
Source: Tracey and Robbins (2006)

Summary. The findings from this study suggest that performance on the ACT is 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).

11.5.4 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 (2012b) 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's degree (Radunzel, 2012), associate's 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 11.3.1). The rates were then summarized across institutions.

Results. As shown in Figure 11.25, 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's 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 (see Appendix A of Radunzel and Noble (2012b) for related figures).

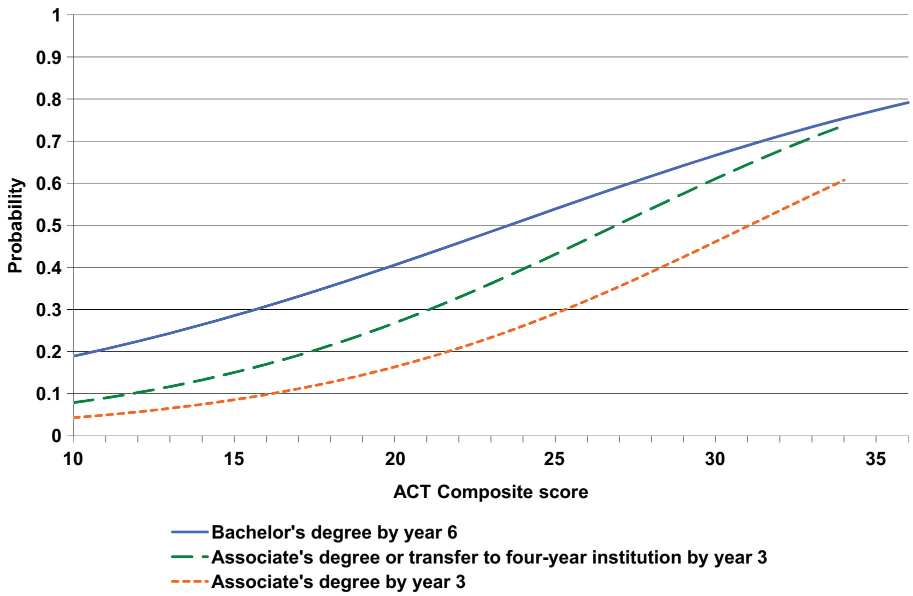


Figure 11.25 Probability of degree completion based on ACT Composite score (Radunzel & Noble, 2012b).

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 11.26 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.

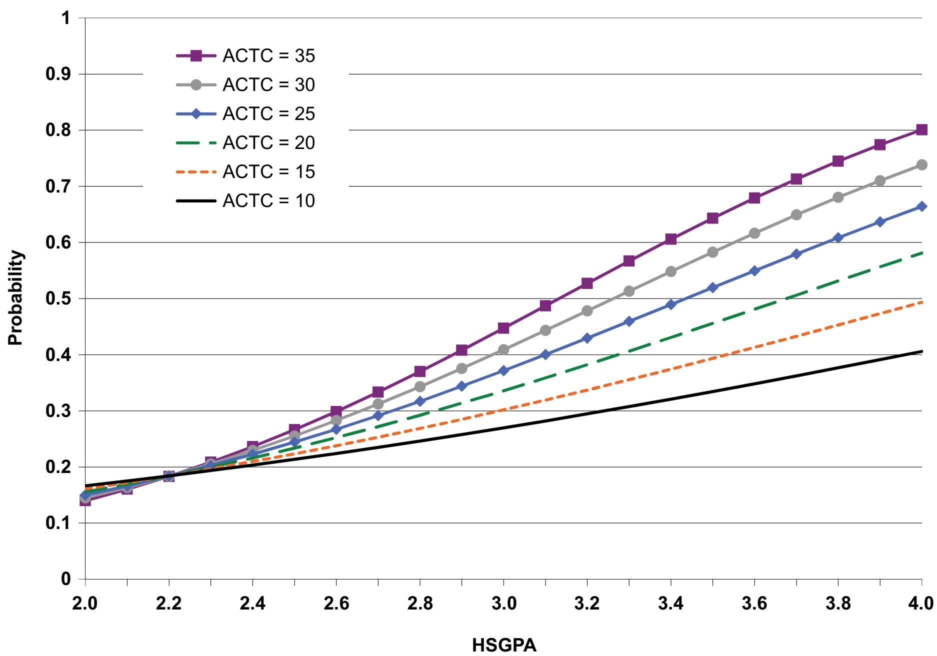


Figure 11.26 Probability of bachelor's degree completion within 6 years, by HSGPA and ACT Composite score (Radunzel & Noble, 2012b).

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 (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, 2012b).

11.5.5 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 of this study are discussed in this section.

Data and method. Longitudinal college outcomes data used in the study were provided by both two- and four-year postsecondary institutions who participated in research services offered by ACT. The study focused on students from the 2005 to 2009 freshman cohorts who declared a STEM major within their first year of college. College outcomes data for the four-year sample were available from 48 four-year institutions and included approximately 53,000 students majoring in STEM who were tracked primarily at the initial institution attended. College outcomes data for the two-year sample were based on more than 10,000 students majoring in STEM who first enrolled in one of 36 two-year institutions from two state systems. For the two-year sample, students were tracked across in-state two- and four-year postsecondary institutions, so in-state transfer information was available.

Due to the nested structure of the data, various hierarchical regression models were used to estimate students' chances of succeeding in a STEM major at a typical institution. Success rates were estimated using the fixed-effect parameter estimates from the hierarchical regression models. Specifically, logistic regression was used to estimate students' chances of earning a cumulative GPA of 3.0 or higher, multinomial regression was used for students' chances of persisting in a STEM major, and discrete-time survival regression was used for students' chances of completing a degree in a STEM-related field. For the four-year sample, completion of a bachelor's degree within 4, 5, or 6 years was evaluated. For the two-year sample, completion of an associate's or bachelor's degree was evaluated. For more details on the data and methods used, see the full report (Radunzel et al., 2015).

Results. ACT STEM scores were positively related to students' chances of achieving specific cumulative GPAs over time, persisting in a STEM major over time, and completing a degree in a STEM field. These findings held for students who began at two- and four-year postsecondary institutions, as well as for students in each of the four STEM major clusters (Computer Science & Mathematics, Engineering, Medical & Health, and Science). Figure 11.27 illustrates results for STEM persistence at years 2, 3, and 4 for students majoring in STEM who began at a four-year institution. The chances of persisting in a STEM major were 67% at year 2, 57% at year 3, and 53% at year 4 for students with an ACT STEM score of 26. Students with STEM scores above 26 had even greater chances of success. In comparison to those with a score of 26 (which represents the ACT STEM Benchmark), students' chances were 12 to 14 percentage points lower across the years for those with an ACT STEM score of 22 (55%, 44%, and 39%, respectively). Moreover, students majoring in STEM with an ACT STEM score of 26 or higher were nearly three times more likely than those with a score of 22 or below to earn a bachelor's degree in a STEM-related field within six years (49% vs. 17%). Only one-third of students majoring in STEM with an ACT STEM score between 23 and 25 earned a bachelor's degree in a STEM-related field by the end of year 6. For additional figures illustrating the relationships between the ACT STEM score and students' chances of succeeding in a STEM major, see the full report (Radunzel et al., 2015).

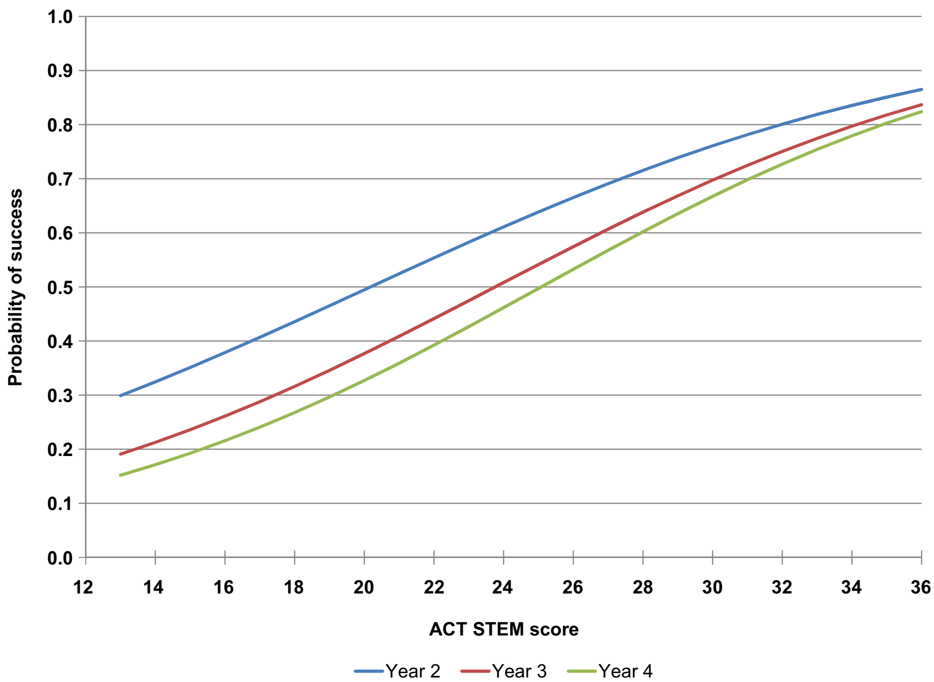


Figure 11.27 Probability of persisting in a STEM major at years 2, 3, and 4 by ACT STEM score at a typical four-year institution.

Summary. The results from the study illustrate 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 hold even after statistically controlling for other student characteristics, such as high school course work taken and grades earned, vocational interests, and demographic characteristics. This finding is consistent with a growing body of literature that has found educational success is a product of not only 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).

11.6 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, 2017) or ACT Aspire, a battery of assessments that measure students' mastery of math, ELA, and science in Grades 3 through 10 (ACT, 2016a).

11.6.1 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 course work, 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 13.

11.6.2 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 benefit 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 that were undertaken to compare the short- and long-term college outcomes between students who had taken dual-credit/dual-enrollment course work 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.

References

(Please note that in 1996 the corporate name "The American College Testing Program" was changed to "ACT".)

ACT. (1997). *ACT Assessment technical manual.* Iowa City, IA: Author.

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

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

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

ACT. (2016a). *ACT Aspire Summative Technical Manual*. Iowa City, IA: Author.

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

ACT. (2017). *PreACT Technical Bulletin*. Iowa City, IA: Author.

Allen, J. D. (2005). Grades as valid measures of academic achievement of classroom learning. *The Clearing House*, *78*(5), 218–223.

Allen, J. (2013). *Updating the ACT College Readiness Benchmarks* (ACT Research Report No. 2013-6). Iowa City, IA: ACT, Inc.

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

Allen, J. (2015b). *Influence of achievement in core high school courses on ACT scores* (ACT Technical Brief). Iowa City, IA: ACT, Inc.

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

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

Allen, J., & Radunzel, J. (2016). *Comparing high school grade point average (GPA) to first-year college GPA* (ACT Data Byte 2016-10). Iowa City, IA: ACT, Inc.

American College Testing Program. (1989). *Using concordance tables.* Iowa City, IA: ACT, Inc.

Andrews, K. M., & Ziomek, R. L. (1998). *Score gains on retesting with the ACT Assessment* (ACT Research Report No. 98-7). Iowa City, IA: ACT, Inc.

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.

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.

Chambers, G. A. (1988). *All of America's children: Variants in ACT test scores—what principals need to know.* Paper presented at the annual convention of the National Association of Secondary School Principals, Anaheim, CA.

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.

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

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: American College Testing Program.

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

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*. (ACT Research Report No. 2016-7). Iowa City, IA: ACT, Inc.

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

Laing, J., & Farmer, M. (1984). *Use of the ACT assessment by examinees with disabilities*. (ACT Research Report Series No. 84). Iowa City, IA: American College Testing Program.

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.

Ling, J., and Radunzel, J. (2017). *Who is likely to graduate high school ready for college?* (ACT Issue Brief). Iowa City, IA: ACT, Inc.

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*. (ACT Research Report No. 2013-8). Iowa City, IA: ACT, Inc.

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*. (ACT Research Report No. 2014-5). Iowa City, IA: ACT, Inc.

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

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®* (ACT Research Report No. 2015-6). Iowa City, IA: ACT, Inc.

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* (ACT Working Paper No. 2015-09). Iowa City, IA: ACT, Inc.

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

Ndum, E., Radunzel, J., & Westrick, P. (2016). *ACT performance for students with disabilities who tested with accommodations in 2014* (ACT Data Byte 2016-16). Iowa City, IA: ACT, Inc.

Noble, J., Crouse, J., Sawyer, R., & Gillespie, M. (1992). *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.

Noble, J., Crouse, J., & Schulz, M. (1995). *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.

Noble, J., Crouse, J., & Schultz, M. (1996). *Differential prediction/impact in course placement for ethnic and gender groups* (ACT Research Report No. 96-8). Iowa City, IA: ACT, Inc.

Noble, J., Davenport, M., & Sawyer, R. (2001). *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 in 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* (ACT Research Report No. 99-4). Iowa City, IA: ACT, Inc.

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* (ACT Research Report No. 99-6). Iowa City, IA: ACT, Inc.

Noble, J., & McNabb, T. (1989). *Differential course work in high school: Implications for performance on the ACT Assessment* (ACT Research Report No. 89-5). Iowa City, IA: American College Testing Program.

Noble, J., Roberts, B., & Sawyer, R. L, (2006*). Student achievement, behavior, perceptions and other factors affecting ACT scores* (ACT Research Report No. 2006-1). Iowa City, IA: ACT, Inc.

Noble, J., & Sawyer, R. (2013). *A study of the effectiveness of developmental courses for improving success in college* (ACT Research Report Series, No. 2013-1). Iowa City, IA: ACT, Inc.

Noble, J., & Sawyer, R. (2002). *Predicting different levels of academic success in college using high school GPA and ACT Composite score*. (ACT Research Report No. 2002-4). Iowa City, IA: ACT, Inc.

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* (ACT Research Report No. 2007-2). Iowa City, IA: ACT, Inc.

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.

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

Radunzel, J. (2012). *Where are 2003 high school graduates seven years later?* Paper presented at the Eighth Annual National Symposium on Student Retention, New Orleans, LA. Retrieved on April 20, 2017 from http://www.act.org/content/dam/act/unsecured/documents/WhereDidHighSchoolGraduatesGo.pdf.

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

Radunzel, J., Mattern, K., & Westrick, P. (2016). *The role of academic preparation and interest on STEM success* (ACT Research Report No. 2016-8). Iowa City, IA: ACT, Inc.

Radunzel, J., & Noble, J. (2012b). *Predicting long-term college success through degree completion using ACT Composite score, ACT Benchmarks, and high school grade point average* (ACT Research Report No. 2012-5). Iowa City, IA: ACT, Inc.

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* (ACT Research Report No. 2013-5). Iowa City, IA: ACT, Inc.

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

Ree, M. J., Earles, J. A., & Teachout, M. S. (1994). Predicting job performance: Not much more than g. *Journal of Applied Psychology*, 79, 518–524.

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* (ACT Research Report No. 2013-4). Iowa City, IA: ACT, Inc.

Sanchez, E. I., & Buddin, R. (2016). *How accurate are self-reported high school courses, course grades, and grade point average?* (ACT Research Report No. 2016-3). Iowa City, IA: ACT, Inc.

Sawyer, R. L. (1985). *Using demographic information in predicting college freshman grades* (ACT Research Report No. 87). Iowa City, IA: American College Testing Program.

Sawyer, R. (2010). *Usefulness of high school average and ACT scores in making college admission decisions* (ACT Research Report No. 2010-2). Iowa City, IA: ACT, Inc.

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* (ACT Research Report No. 2013-9). Iowa City, IA: ACT, Inc.

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, Inc.

Schiel, J., & Harmston, M. (2000). *Validating two-stage course placement systems when data are truncated* (ACT Research Report No. 2000-3). Iowa City, IA: ACT, Inc.

Schiel, J., & Noble, J. (1992). *The effects of data truncation on estimated validity indices for course placement* (ACT Research Report No. 92-3). Iowa City, IA: American College Testing Program.

Schiel, J., Pommerich, M., & Noble, J. P. (1996). *Factors associated with longitudinal educational achievement, as measured by PLAN and ACT Assessment scores* (ACT Research Report Series No. 96-5). Iowa City, IA: American College Testing Program.

Schiel, J. L., & Valiga, M. J. (2014a). *How ACT scores change with test preparation* (ACT Information Brief 2014-6). Iowa City, IA: ACT, Inc.

Schiel, J. L., & Valiga, M. J. (2014b). *How the amount of time spent on test preparation relates to ACT scores* (ACT Information Brief 2014-7). Iowa City, IA: ACT, Inc.

Schiel, J. L., & Valiga, M. J. (2014c). *How particular outcomes of test preparation relate to ACT scores* (ACT Information Brief 2014-8). Iowa City, IA: ACT, Inc.

Scholes, R. J., & Lain, M. M. (1997). *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). *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.

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.*

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.

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, Office of Educational Research and Improvement. (1994). *What do student grades mean? Differences across schools* (Office of Research Report 94-3401). Washington, DC: Author. Accessed at https://www2.ed.gov/pubs/OR/ResearchRpts/grades.html.

Valiga, M. (1986). *Accuracy of self-reported high school course and grade information* (ACT Research Report 87-1). Iowa City, IA: American College Testing Program.

Westrick, P. A. (in press). Reliability estimates for undergraduate GPA*. Educational Assessment*.

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

Westrick, P., & Allen, J. (2014). *Validity evidence for ACT Compass placement tests* (ACT Research Report No. 2014-2). Iowa City, IA: ACT, Inc.

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.

Whitney, D. R. (1989). Educational admissions and placement. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 515–525). New York: 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* (ACT Research Report 2004-4). Iowa City, IA: ACT, Inc.

Woodruff, D. J., & Ziomek, R. L. (2004b). *Differential grading standards among high schools* (ACT Research Report 2004-2). Iowa City, IA: ACT, Inc.

Zhang, Q., & Sanchez, E. I. (2013). *High school grade inflation from 2004 to 2011* (ACT Research Report No. 2013-3). Iowa City, IA: ACT, Inc.

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* (ACT Research Report No. 96-7). Iowa City, IA: ACT, Inc.

Ziomek, R. L., & Svec, J. C. (1995). *High school grades and achievement: Evidence of grade inflation* (ACT Research Report 953). Iowa City, IA: ACT, Inc.

Chapter 12

Online Testing and Mode Comparability

12.1 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 offered to all state and district sites in 2016 and will continue to be offered going forward.

Today, the ACT may be administered on paper or online. The ACT administered online is the same test as the paper version but presented in an online delivery format. Online testing of the ACT is delivered in testing windows, which are designed to provide test access over a short period of time and to accommodate makeup and emergency situations. Online administration of the ACT follows the administration guidelines established for paper testing, where appropriate.

12.2 Online Platform and Capabilities

ACT collaborated with Pearson to design the platform architecture for administering the ACT online test delivery system. Test centers can use this test delivery system across multiple forms of technology (operating systems, platforms, and devices), including personal computers such as Macintosh® or Microsoft Windows. 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

12.3 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 studies. Initial online forms were linked to paper forms through equating methodologies based on data gathered in special mode comparability studies where both paper and online forms were administered. Subsequent online forms are equated to the online base forms through online test equating studies. ACT uses the same data collection designs and test equating procedures to link online scores to paper scores and to equate the online forms as it uses to equate the ACT paper test forms.

12.4 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 ensure 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.

All three studies used a randomly equivalent groups design. 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 both 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 second mode study. Below are brief summaries of these studies. See Li, Yi, and Harris (2017) for more details.

12.4.1 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 the three timing conditions: the current standard paper time limit (i.e., 45, 60, 35, and 35 minutes for English, mathematics, reading, and science tests, respectively), the current time limit plus five minutes, and the current time limit plus ten 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 extracted 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 paper timing condition might be more speeded than paper testing. For example, compared with the matched operational paper sample, the average number of items omitted was higher for the timing study sample for all subject tests under the current standard paper testing timing condition. The timing study sample also had lower item *p*-values for the last few items than the matched sample, especially for reading and science. In addition, among the students who responded to the survey questions, about half either disagreed or strongly disagreed with the statement that they had enough time to complete the reading and the science tests.

However, findings from the timing study might have been confounded with issues of low motivation and unfamiliarity with the online testing format. For example, even though an online tutorial was provided for students to view 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 results of various analyses were evaluated from different perspectives, ACT decided to tentatively increase online testing time for the reading and science tests by five minutes. Also, ACT planned a subsequent mode comparability study to continue evaluating the timing issue.

12.4.2 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 where 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 (one paper and two online) that were administered in the study. The assignment was similar to distributing spiraled paper booklets. After the administration, survey questions were sent to students who participated in the study for collecting their comments and feedback on their testing experience.

More than 7,000 students from about 80 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 examination of the random assignment. Students with invalid scores and test centers with large discrepancies in form counts across modes were excluded from further analyses.

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 showed that although little difference was found between the two modes in terms of test reliability, correlations among tests, effective weights, and factor structures, item scores and test scores tended to be higher and omission rates tended to be lower for the online group than for the paper group, especially for the reading and science tests. Equating methodology was used for all four multiple-choice tests to adjust for mode differences to ensure that the college reportable scores of students participating in the mode comparability study were comparable to national test takers, regardless of the testing mode.

Based on the findings from the spring 2014 mode comparability study, ACT decided to eliminate the extra five 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.

12.4.3 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 five 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. Since the spring 2015 study followed the same design as the 2014 study, 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 five minutes. 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.

12.4.4 Summary

The ACT online timing study and the two mode comparability studies all used a solid research design involving random assignment of examinees 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 high.

Whereas the analyses showed no evidence of differences in the measurement of the construct or in measurement precision, slight differences were found on item level and test level statistics. Under the final online timing conditions, the largest mean between-mode difference was found for the reading test, which was about one scale score point (with an effect size of 0.18). Considering that the standard error of measurement of the test is about two scale score points, the mode difference is 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 strict comparability of scores between paper and online administrations. Subsequent online test forms are equated to the base online form, which has been linked to paper forms through the mode study, to ensure that scores from the ACT test forms are all comparable regardless of mode.

Reference

Li, D., Yi, Q., & Harris, D. (2017). *Evidence for paper and online ACT® comparability: Spring 2014 and 2015 mode comparability studies* (ACT Research Report No. 2017-1). Iowa City, IA: ACT, Inc.

Chapter 13

Growth Models Using ACT Test Scores

13.1 Overview

Understanding student growth models can help students, parents, educators, and practitioners make better use of ACT data. Growth models can be used to answer important questions such as: How does the growth of students from my school compare to national growth averages? How much does my student need to grow to reach her or his ACT score goal? 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 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.

13.2 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 & 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.

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, where will she or he 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 and residual gain score models are supported by ACT's Growth Modeling Resources, as discussed later in this chapter.

Reliance on common score scales. Gain-based models require that test scores from multiple time points share a common scale. This can be achieved by 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.

13.3 ACT Growth Modeling Resources

To help users implement growth models based on ACT assessments, ACT provides normative growth data and support for conditional status models for various assessment combinations and grade levels. ACT's Growth Modeling Resources (http://www.act.org/content/act/en/research/act-growth-modeling-resources.html) include support for the SGP model, obtaining score projections, and the residual gain score model. Examples are provided for these models, including steps for aggregating SGPs and residual gain scores to support value-added interpretations.

13.3.1 Student Growth Percentile Model

The Student Growth Percentile (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 a SGP of 75 performed as well as or better than 75 percent of her or his academic peers with similar score histories. SGPs supported 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 R statistical software package (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., race/ethnicity, gender, and economically disadvantaged status). Measures of aggregate growth include the mean and median SGP. Recent 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 economically disadvantaged students might be expected to have lower mean SGP than a school serving students from affluent 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. SGP lookup tables available for the ACT test include:

- ACT Aspire-to-ACT. The reference group consists of examinees who took ACT Aspire in spring Grade 10 and the ACT test in spring Grade 11 in consecutive years (one year apart) from spring 2013 through spring 2016.
- ACT-to-ACT. The reference group consists of examinees who took the ACT test in Grades 11 and 12 (6 months apart) in consecutive years from 2013 through 2016.
- ACT Plan-to-ACT. The reference group consists of examinees who took ACT Plan in Grade 10 and the ACT test in Grade 11 approximately one and a half years apart in consecutive years from 2006 through 2016.

When available, SGP lookup tables will also be provided for PreACT to the ACT test. SGPs are currently provided for English, mathematics, reading, and science. SGPs are also provided for writing where available (the writing test was not available for ACT Plan and is not available for PreACT). The ACT Growth Modeling Resources website also provides examples of how to apply the SGP model.

13.3.2 Projection and Residual Gain Score Models

The projection model is primarily used to predict future test scores from current and past test scores. It is meant to answer the question “Given this student’s observed current and past scores, and based on patterns of scores in the past, where is this student likely to score in the future?” (Castellano & Ho, 2013). Predicted ACT scores can be compared against a target score, which could be a future grade’s proficiency cut score (e.g., ACT College Readiness Benchmark) or a goal tailored for each student. Students can be considered “on target” for meeting their goal if their predicted score is greater than or equal to the goal score.

The projection model supported by ACT uses linear regression to establish an equation relating students’ current and past scores to their future scores. The projection model is flexible in that multiple current and past scores (in the same subject or from different subjects), as well as other measures, can be used to predict future scores.

For example, Grade 11 ACT mathematics score can be predicted based on Grade 10 ACT Plan scores in all four subject areas (English, mathematics, reading, and science) and on the number of months between the two assessments:

> Predicted ACT Mathematics Score = $\beta 0$ + $\beta 1$ × ACT Plan English Score + $\beta 2$ × ACT Plan Mathematics Score + $\beta 3$ × ACT Plan Reading Score + $\beta 4$ × ACT Plan Science Score + $\beta 5$ × Months Elapsed

In this model, the β values are weights relating each prior test score to the future test score. These weights are referred to as projection parameters. Predicted ACT mathematics score is determined by ACT Plan scores in all four subject areas, as well as the number of months elapsed between the ACT Plan and ACT tests. Prediction equations that are available from the ACT Growth Modeling Resources take a similar form.

The projection model relies on regression assumptions, such as normally-distributed error terms with constant variance. Projection models can make predictions multiple years into the future and can use more than one year of current or prior test scores (predictors). Currently, the projection models supported by the ACT Growth Modeling Resources only use one year of test scores.

The residual gain score model can be used in conjunction with the projection model. The projection model produces an expectation for the current year score based on past score(s). The residual gain model describes the difference between the actual score and the expected score. This difference (actual score – expected score) is called a “residual” in the context of regression and a “residual gain score” in the context of the residual gain score model. Similar to the SGP model, the residual gain score model describes growth in a normative fashion. The sample used to estimate the projection parameters is the reference group.

The ACT Growth Modeling Resources includes projection parameters for several pairs of assessments, including some that can be used to predict ACT scores:

- ACT Explore-to-ACT. Examinees who took ACT Explore in Grade 8 and then took the ACT test in Grade 11 (27 to 45 months apart).
- ACT Plan-to-ACT. Examinees who took ACT Plan in Grade 10 and then took the ACT test in Grade 11 (10 to 14 months apart).
- ACT Plan-to-ACT. Examinees who took ACT Plan in fall Grade 10 and then took the ACT test in spring Grade 11 (15 to 21 months apart).

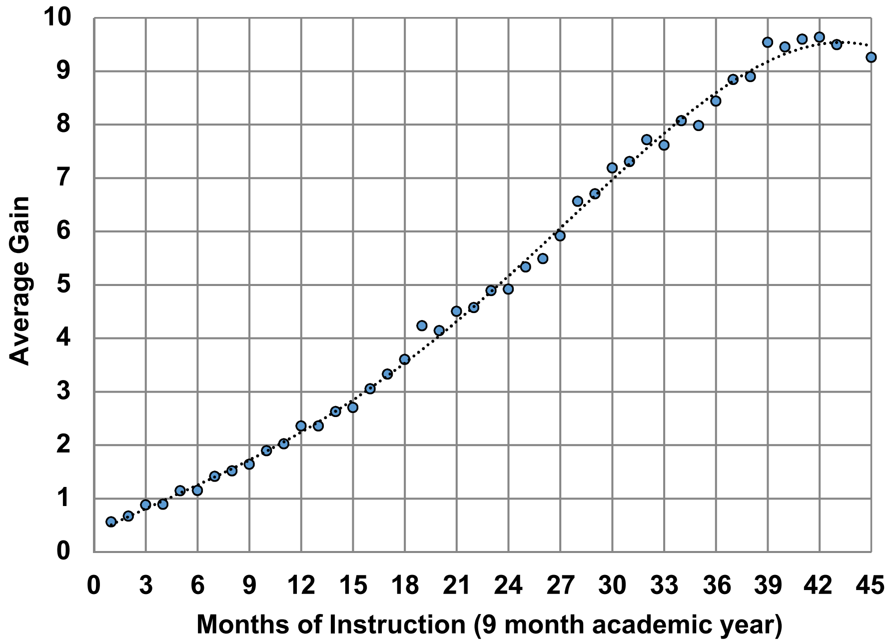
Projection parameters for all pairs of assessments are provided for four subject areas: English, mathematics, reading, and science. The prediction equations include prior test scores in four subject areas and the number of months elapsed between the two assessments. The growth modeling resources also include documentation of how to apply the projection model and examples of how to produce residual gain scores.

13.4 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 course work and grades. This section summarizes research that explains some of the variation in student growth.

13.4.1 ACT Score Gains by Months of Instruction

Camara and Allen (2017) examined the relationship of 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 captures 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. They found that ACT scores steadily increase with more instructional time (Figure 13.1). 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.



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

Figure 13.1 Average gain in ACT Composite score, by months of instruction

13.4.2 Predictors of Long-Term Growth

13.4.2.1 Benefits of Additional High School Course Work 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 extending more effort in these courses. Sawyer (2008) examined the effectiveness of taking additional courses and earning higher grades for improving high school students' academic preparation for college, using data from students who took ACT Explore in eighth grade, ACT Plan in tenth grade, and the ACT in eleventh or twelfth grade.

Data. The sample included students who took all three tests (ACT Explore, ACT Plan, and the ACT) and graduated from high school in 2005 or 2006. The source data set for 2005 contained records for 98,812 students from 4,191 high schools. The source data set for 2006 contained records for 117,280 students from 4,638 high schools. Data for the 2005 cohort were used to select variables for predicting ACT English, mathematics, reading, and science scores. The same model was then re-estimated using the 2006 cohort.

ACT Explore scores in all four subject areas were used as measures of prior achievement. Other predictors included students' high school, background characteristics, ACT testing characteristics, standard high school course work taken and average course grades, and advanced, accelerated, or honors courses taken. The following background variables were used as predictors: gender, race/ethnicity, parents' educational level, family income, English as the primary language spoken at home, and the state in which a student's high school was located. ACT testing characteristics included students' age and grade level at the time of ACT testing and retesting status.

Method. Hierarchical linear models were used to relate the predictors to ACT test scores. Random-intercept main-effects models were estimated, in which all regression coefficients, except for the intercept, were constrained to be the same for each high school. Additionally, to examine the variability of regression weight across high schools, random effect for each predictor in the main-effects model was estimated. Missing values in the predictor variables were imputed. Interactions were considered, including age at time of testing by grade level at time of testing and retesting dummy variables; grade level at time of testing by retesting dummy variables; ACT Explore score by standard course work, advanced or honors course work (Adv./Hon.), and course grade averages; and standard course work and Adv./Hon. course work by course grade averages. School means of the predictor variables, as well as two state dummy variables corresponding to statewide testing in Colorado and Illinois, were included as potential Level-2 predictors of the intercept.

Aggregate Benefit of Additional Course Work and Higher Grades. To estimate the benefit of taking additional courses and earning higher grades, the percentage of students who would meet the various ACT College Readiness Benchmarks under each of the following "high effort" scenarios was calculated: 1) with current Explore scores, course work, and grades, 2) increasing Explore scores by two points in each subject area, 3) taking one additional standard college-preparatory course of each type present in the model, 4) taking Adv./Hon. courses in each relevant subject area, and 5) increasing the grade average in each relevant subject area by one letter grade. In setting up the scenarios, the value of each predictor variable was capped at its maximum: for example, a student who already had a 4.0 grade average could not increase her or his grade average. Next, for each student, a predicted ACT score was calculated using the relevant hierarchical regression model. A random error term was added to each predicted ACT score representing the residual variation of actual ACT scores around the predicted ACT scores; the resulting quantity was a simulated ACT score. Finally, the percentage of the simulated ACT scores that met or exceeded the relevant College Readiness Benchmark were calculated.

Alternative "moderate effort" scenarios included 1) meeting the ACT Explore Benchmark scores in each subject area, 2) taking the minimum recommended standard college-preparatory courses in the subject areas relevant to the model, and 3) earning a B or higher grade average in each subject area relevant to the model. The original report includes more details on the methods (Sawyer, 2008).

Results. In predicting ACT English score and ACT reading score, neither English standard course work nor social studies standard course work were statistically significant (p <.001), either for the 2005 cohort or 2006 cohort. Foreign language course work was statistically significant (p <.001) in predicting the ACT English score but not the ACT reading score. In contrast, both mathematics standard course work and science standard course work were statistically significant (p <.001) in predicting both the ACT mathematics and ACT science scores. Adv./Hon. Course work in English and Adv./Hon. Course work in social studies were statistically significant (p <.001) in predicting both the ACT English and ACT reading scores. Adv./Hon. Course work in mathematics and Adv./Hon. Course work in science were also statistically significant (p <.001) in predicting both the ACT mathematics and ACT science scores. Similarly, grade averages in English and social studies were statistically significant (p <.001) in predicting both the ACT English and ACT reading scores. Finally, grade averages in mathematics and science were statistically significant (p <.001) in predicting both the ACT mathematics and ACT science scores.

All four testing context variables were statistically significant (p <.001) in all models except for ACT mathematics. The estimated coefficients indicate that: students who test at an older age tend to score lower than students who test at a younger age (by 0.4 to 0.8 score points per year, depending on subject area); students of a given age who test in grade twelve tend to score higher on the ACT English, reading, and science tests than students of the same age who test in grade eleven (by 0.4 to 0.9 score points); students who retest and update their course work and course grade information tend to score higher than similar students who test only once (by 0.2 to 1.0 score points); and students who retest, but do not update their course work and course grade information tend to score lower than similar students who test only once (by 0.4 to 0.9 score units).

Relative Importance of Predictor Variables. The standardized coefficient "beta weight" results indicate that for all four ACT test scores, prior educational achievement (as measured by ACT Explore scores) was much more important than any other class of predictor variables, including standard course work. Prior achievement was more strongly related to ACT English and reading scores (beta weight sum = 0.74 and 0.72, respectively) than to ACT mathematics and science scores (beta weight sum = 0.54 and 0.64, respectively). Given that ACT Explore scores likely also affect ACT scores indirectly through course work and grades, the total effects of ACT Explore scores could even be larger.

The beta weight results also suggest that taking standard and Adv./Hon. mathematics and science courses improves ACT mathematics and science scores but that taking English, social studies, and foreign language courses is of little or no benefit in improving ACT English and reading scores. Earning higher grades in standard courses and taking Adv./Hon. courses provide modest benefit. Given the strong relationship between ACT Explore scores and ACT English and reading scores, however, major improvements in reading and writing need to occur *before* eighth grade.

Interaction Models. Coefficients of interaction terms all indicate that students with high ACT Explore scores benefit more from standard or Adv./Hon. course work than do students with low ACT Explore scores. Table 13.1 shows the expected increase in ACT reading score from taking an Adv./Hon. social studies course or from increasing social studies grade average. Note that the expected increase resulting from either form of enhanced preparation depends on the ACT Explore reading score. The models also include interaction terms for course work by subject area grade average, indicating that students who earn high grades in particular subject areas benefit more from taking courses in those areas than do students with low grade averages.

Simulation Study Results. Among the forms of enhanced preparation, increasing ACT Explore scores would result in the greatest increase in the percentage of students meeting the ACT College Readiness Benchmarks. Increasing ACT Explore scores by two points yields increases of 12, 13, 16, and 13 percentage points in meeting the English, mathematics, reading, and science Benchmarks, respectively. Taking additional standard courses, taking Adv./Hon. Courses, and earning higher grades would result in only a modest increase in the percentage of students meeting the ACT Benchmarks.

Summary. Students' background characteristics, ACT Explore scores, high school attended, high school course work, and high school grades were all related to ACT scores, but ACT Explore scores were by far the most strongly related. Improving ACT Explore scores was likely to be more effective in improving

ACT scores than other forms of enhanced preparation. Taking more standard or advanced courses in high school and earning higher grades was more beneficial to students who had high ACT Explore scores to begin with. There was significant variation in high schools' average ACT scores, even after accounting for differences in their students' characteristics. The benefit of additional standard course work, advanced/honors course work, and higher grades also varied significantly among high schools.

Table 13.1. Expected Increase in ACT Reading Score from Enhanced Preparation, Given ACT Explore Reading Score

	Enhanced preparation	
ACT Explore Reading Score	Take Adv./Hon. social studies course	Raise social studies grade average one letter grade
25	0.89	0.96
20	0.70	0.71
15	0.51	0.45
10	0.33	0.19

Bassiri (2014) replicated Sawyer's study using a more recent cohort of students (high school graduates of 2013 who took ACT Explore in eighth 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 course work in English and course work in social studies being significant predictors of ACT English and ACT reading scores, respectively.

13.4.2.2 Predictors of Academic Growth in Secondary School among Academically Advanced Youth

Many academically advanced youth take the ACT test in seventh grade for academic talent searches and again in eleventh or twelfth grades for college admissions (Allen, 2016), enabling an investigation of predictors of growth during secondary school. Wai and Allen (2017) tested whether variation in academic growth among academically advanced youth is explained by socio-demographics, high school characteristics, course work taken, high school GPA, Holland-type vocational interests (Holland, 1997) or extracurricular activities.

Data. The sample consisted of over 460,000 students who took the ACT test in seventh grade and again in eleventh or twelfth grade and were projected to complete high school between 1996 and 2016. The vast majority of students in the sample (96%) sent their ACT score results to a major talent search program when in seventh grade.

Method. Academic growth was measured using the residual gain model. Last ACT Composite score obtained in high school was regressed on seventh grade ACT Composite score, the number of months between the two tests, and indicators for whether students tested with special accommodations. Predictors considered in the study included socio-demographic variables, ACT Interest Inventory[1] scores, high school characteristics, high school course work and GPA, and extracurricular activities. Missing predictor data were imputed.

Multiple regression was used to relate the full set of predictor variables to academic growth. To facilitate interpretations, the continuous variables (residual gain ACT Composite score, school class size, high school GPA, and ACT Interest Inventory scores) were standardized to have a mean of 0 and standard deviation of 1. All categorical variables (school type, school locale, race/ethnicity, income level, parents' educational level, elective and advanced course work, and extracurricular activities) were dummy-coded. Parents' educational level was only collected for students in cohorts of 2011–2016.

Results. Due to the large sample size, most predictors were statistically significant, even if the regression coefficient was very small. Overall, the model accounted for 29% of the explainable variance in academic growth. Relative to white students, African American ($\beta = -0.377$), Hispanic ($\beta = -0.198$), and students of other minority groups ($\beta = -0.069$) had lower academic growth. Also, low-income ($\beta = -0.172$) and middle-income ($\beta = -0.092$) students had lower growth than high-income students, and males ($\beta = 0.249$) had higher growth than females. Academic growth varied by type of high school in the following order: Catholic, private, low-poverty public, moderate-poverty public, home school, and high-poverty public.

Students with higher high school GPA demonstrated higher growth ($\beta = 0.226$). The elective high school course with the strongest positive relationships with academic growth was Calculus ($\beta = 0.134$), followed by Trigonometry, Chemistry, Physics, other math beyond Algebra 2, and other foreign language. Elective social studies courses (Geography, Psychology, Economics, and other history) and courses in the arts had negative relationships. Students taking advanced course work (advanced, accelerated, or honors courses) had significantly higher growth. The effect was strongest for advanced mathematics ($\beta = 0.109$), followed by social studies ($\beta = 0.090$), and natural sciences ($\beta = 0.062$). Higher Science & Technology vocational interest scores (corresponding to the *Investigative* personality type) were related to higher growth ($\beta = 0.078$), as were higher Business Operations scores (*Conventional* personality type) ($\beta = 0.043$). Higher Technical scores (*Realistic* personality type) ($\beta = -0.040$) were related to lower growth. Relative to other predictors, extracurricular activities were not as predictive of academic growth.

[1] In Holland's theory of vocational interests/choices (1997), both individuals and environments can be represented by six personality types. The ACT Interest Inventory scales and the corresponding Holland types (in parentheses) are: Science & Technology (Investigative), Arts (Artistic), Social Service (Social), Administration & Sales (Enterprising), Business Operations (Conventional), and Technical (Realistic).

Summary. The study examined predictors of student growth over a long period of time: seventh grade to eleventh or twelfth grade. Predictors of growth included malleable factors such as high school course work and grades, and background variables such as race/ethnicity and family income. All variables combined—socio-demographics, interests, high school characteristics, high school course work and GPA, and extracurricular activities—explained 29% 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 home-schooled students and students attending high-poverty public schools showed 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.

13.4.3 Subgroup Differences in Growth

13.4.3.1 Academic Growth Patterns for English Language Learners and Students with Disabilities

Bassiri and Allen (2012) examined differences in growth for English language learners (ELLs) and students with disabilities (SWD).

Data. This study used longitudinal data on 103,725 students who took ACT Explore in eighth grade, ACT Plan in tenth grade, and the ACT test in eleventh or twelfth grade. As part of the ACT Explore and ACT Plan programs, some schools provided extra demographic and subgroup data, including indicators for which students were classified as ELL and which are in special education (SPED) programs. Because the school-reported SPED indicator is not demarcated by type of disability, the student-reported disability subgroups were included to examine growth differences for students with different types of disabilities. ELLs made up 2% of the sample, students in SPED programs made up 7%, and SWD made up 6% of the sample. Of the SWD, 60% had a cognitive/learning disability, 17% had a physical disability, and the other 23% were classified as having some other type of disability.

Method. Seven groups of students were identified for analysis: 1) a reference group that included students who were not classified as ELL, SPED, or SWD, 2) ELL, 3) SPED, 4) SWD, 5) physical disability, 6) cognitive/learning disability, and 7) other type of disability. Note that the ELL, SPED, and SWD groups are not mutually exclusive–students could be members of more than one of these groups. Also, groups 5–7 are subsets of group 4.

Academic growth was measured using the residual gain score model. ACT Plan (ACT) scores were regressed on prior ACT Explore (Plan) scores in all four subject areas, the number of months that elapsed between the two tests, and the prior subject-specific mean ACT Explore (Plan) score for the high school using a two-level hierarchical linear regression model (Raudenbush & Bryk, 2002) with random intercepts. The residual scores were aggregated by subgroups to form measures of aggregate growth for each subgroup. This was done separately for each subject area (English, mathematics, reading, and science), and for each grade-level span (8 to 10, 10 to 11/12, and 8 to 11/12). In all, twelve models were fit.

Results. ***Grade 8 to 10.*** ELLs were the only group that had statistically significant (p-value < 0.01) and positive residual gain scores in mathematics (0.43) and science (0.28) from Grade 8 to 10. The SPED and SWD groups had negative residual gain scores in mathematics and science. Less-than-expected growth in English and reading was observed in all subgroups except in reading for students with physical disabilities.

Grade 10 to 11/12. During this period, the SPED, physical disability, and other disability groups had less-than-expected growth in all four subject areas. ELL students continued to have above-average growth in mathematics (0.15) but below-average growth in science (−0.44). Students with cognitive/learning disabilities experienced above-average growth in all subject areas, particularly in reading (0.38) and science (0.21).

Grade 8 to 11/12. Once again, ELL students continued to show above-average growth in mathematics (0.68). The least growth was observed for SPED students in English (−0.92). In general, the residual gain scores for this grade span were statistically significant if one or both residual gain scores were significant for the shorter spans (Grade 8 to 10 and/or Grade 10 to 11/12).

Summary. Across subject areas, the average ACT Explore, ACT Plan, and ACT scores were largest for the reference group and smallest for the ELL and SPED groups. Among SWD, the average scores across subject areas were lower for students with cognitive/learning disabilities than those for students with physical disabilities. In some cases, the growth measures revealed a different pattern. For example, compared to the reference group, ELL students had consistently higher growth in mathematics and science between Grade 8 and Grade 10; SWD experienced 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 grew 0.28 points less-than-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 for Explore, 21.6 for the ACT). The 0.28 score point growth difference is only about 5% (0.28 / 5.10) of the reference group's overall gain and so is not very large 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 grew, on average, 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, SPED students grew nearly a full point less in English and students in the other disability subgroup grew over half a point less in mathematics.

13.4.3.2 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 to 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.

Data and method. This study used longitudinal data for approximately 282,000 students who took ACT Explore in eighth grade, ACT Plan in tenth grade, and the ACT test in eleventh or twelfth 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 are also referred to as non-first-generation.

Of the original 281,854 students, 56,162 were first-generation students (20%), 187,712 were non-first-generation students (67%), and 13% did not report their parents' educational level. Students who did not report their parents' educational level were excluded from the study. Of non-first-generation students, 41% of students had at least one parent with some college experience, 35% with at least one parent holding a bachelor's degree, and 24% with at least one parent holding a graduate degree.

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 (8 to 10, 10 to 11/12, and 8 to 11/12).

Results. Figure 13.2 displays the mean residual scores from ACT Explore to the ACT by subgroup and 99% confidence intervals (error bands) around each mean residual score. Similar trends were observed for the other two growth periods. Across subject areas, the average ACT Explore, ACT Plan, and ACT scores increased with parents' educational level. For example, the average ACT score difference between first- and non-first-generation students whose parents had some college experience, bachelor's degree, or graduate degree were, respectively, 1.0, 2.3, and 3.2 in English; 0.8, 2.0, and 2.8 in mathematics; 0.9, 2.1, and 2.9 in reading; and 0.8, 1.9, and 2.5 in science. The 2013 average ACT scores nationwide in English, mathematics, reading, and science were 20.2, 20.9, 21.1, and 20.7, respectively (ACT, 2013). The average ACT scores for first-generation students were, respectively, 0.35, 0.64, 0.54, and 0.43 standard deviations lower than those of national ACT scores. On the other hand, the corresponding averages for non-first-generation students whose parents had at least a bachelor's degree were, respectively, 0.52, 0.24, 0.33, and 0.45 standard deviations higher than those of national ACT scores.

Across all grade-level periods and subject areas, the multiple correlations for the models ranged from 0.76 to 0.87 indicating strong relationships between predicted and actual later test scores. First-generation students across all grade level periods experienced less-than-expected growth in all subject areas, ranging from −0.52 to −0.18 in English; −0.39 to −0.15 in reading; −0.34 to −0.19 in mathematics; and −0.35 to −0.15 in science. Similarly, students whose parents had only some college education experienced less-than-expected growth in all subject areas and across all grade periods. However, students whose parents had at least a bachelor's degree had statistically significant (p-value < 0.01) and positive mean residual scores (ranging from 0.09 to 0.79 in English, 0.08 to 0.63 in reading, 0.13 to 0.71 in mathematics, and 0.08 to 0.53 in science).

In general, growth in English and reading tended to be higher for females than males across all parents' educational levels. That is, females and males both had negative residual values at lower parents' educational levels, but females still had higher growth values as compared to males (with the exception of 8 to 11/12 growth in reading for males). Male students consistently experienced at or above-average expected growth in mathematics and science while female students typically experienced less-than-expected growth in both subject areas. The only exception was in mathematics between Grades 10 to 11/12 and 8 to 11/12 where female students whose parent had a graduate degree had above-average growth.

Summary. Across subject areas, the average ACT Explore, ACT Plan, and ACT scores increased with parents' education level. First-generation students and students whose parents had some college experience both had statistically significant (p-value < 0.01) 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 become more pronounced over time. That is, the mean residuals are larger in later grades, indicating that educational disparity by socioeconomic status is exacerbated over time. Future research should examine potential causes of the growth differences, such as low-income status, type of high school course work, and high school grades.

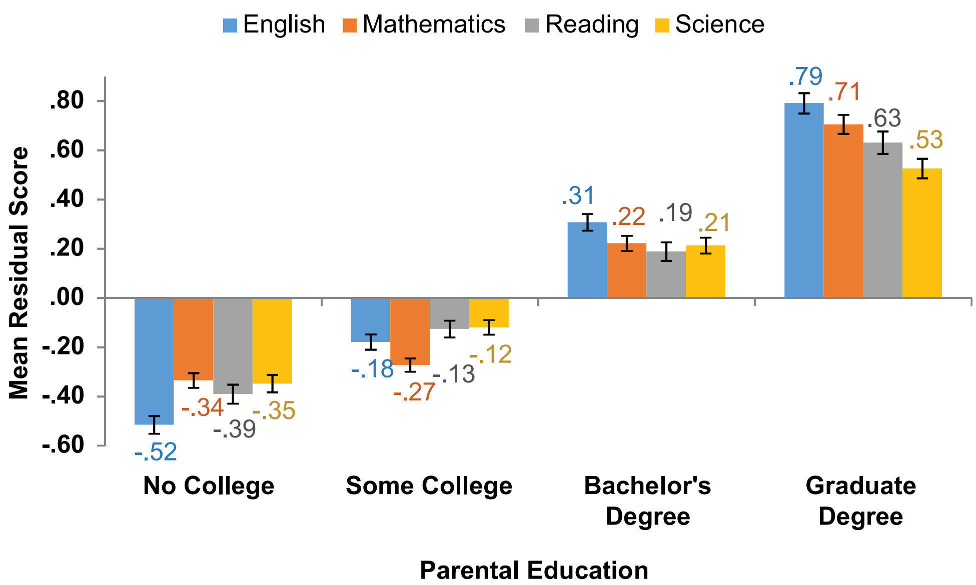


Figure 13.2 Mean residual scores for Grade 8 to 11/12 growth period by parental education.

13.5 Using Growth Models for Evaluation of Programs and School Effectiveness

13.5.1 Example of Program Evaluations

The federal government's Gaining Early Awareness and Readiness for Undergraduate Program (GEAR UP) is designed to increase the number of low-income students who are 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, 2007a).

Data. The sample consisted of two cohorts of students from GEAR UP schools: students that took ACT Explore in Grade 8 during the 2002–2003 academic year and later took ACT Plan in Grade 10 during the 2004–2005 academic year, and a second cohort of students that took ACT Explore in

Grade 8 during the 2003–2004 academic year and later took ACT Plan in Grade 10 during the 2005–2006 academic year. For each GEAR UP school that participated in ACT Explore and ACT Plan, a Non–GEAR UP school was selected that also participated in ACT Explore and ACT Plan to serve as a comparison group. Each Non–GEAR UP school was matched to a particular GEAR UP school based on school achievement and other school characteristics (Table 13.2). The main criterion for matching schools was to keep the difference in mean ACT Explore Composite score within one point. The matching process was executed in the same fashion for both cohorts. For more details on the matched data, see the Appendix in the full report that includes tables that compare the GEAR UP and Non–GEAR UP groups on school and student-level characteristics (ACT, 2007a).

Table 13.2. Sample Sizes for Cohorts Studied

Cohort	Year and assessments	Number of matching school pairs	Number of students	
			GEAR UP	Non–GEAR UP
1	2002–2003 Explore & 2004–2005 Plan	119	6,270	5,808
2	2003–2004 Explore & 2005–2006 Plan	136	6,707	5,791

Method. The focus of study was on the college preparedness and college intent of students in GEAR UP schools as compared to their matched counterparts (Non–GEAR UP schools). The outcome variables included changes in ACT Explore and ACT Plan Composite scores, meeting ACT Explore and ACT Plan College Readiness Benchmarks for each subject area, plans for taking core high school curriculum at Grade 10, and changes in plans for college from Grade 8 to Grade 10. For all the outcomes except for the first, hierarchical logistic regression models were used with random intercept models. For the first outcome (change in Composite score from Explore to Plan) hierarchical linear regression models were used. To control for the discrepancy in poverty level between GEAR UP schools and their matched Non–GEAR UP counterparts, school poverty level (proportion of students eligible for free or reduced lunch) was included as a covariate.

Results. In general, analyses suggest that students from GEAR UP schools had similar to slightly greater changes in overall academic performance from Grade 8 to Grade 10 as compared to the Non–GEAR UP comparison group, after adjusting for the school's poverty level. Specifically, for one cohort, students in the GEAR UP group gained 0.16 more composite scale score points as compared to Non–GEAR UP students (Table 13.3). Additionally, the odds of being college-ready were 16% and 27% higher for the GEAR UP group in English and reading, respectively. Students from GEAR UP schools were also slightly more likely to take the core high school curriculum and have plans for college at Grade 10. However, for the other cohort, there was no significant difference in overall academic performance, in taking the core high school curriculum, or having plans for college.

Table 13.3. Changes in Mean Composite Score by GEAR UP group[2]

Mean Composite Score	GEAR UP	Non–GEAR UP
EXPLORE (Grade 8)	14.53	14.56
PLAN (Grade 10)	16.36	16.32
Change	+1.83	+1.76
GEAR UP increase	+0.07	
After adjustment for poverty level	+0.16	

Summary. In general, analyses suggest positive GEAR UP effects, though the effect sizes were generally small and the 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 with the research design. For more details on the study's limitations and recommendations, see the full report (ACT, 2007a).

13.5.2 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.

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

Allen, Bassiri, and Nobel (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.

Data. The sample consisted of 485 high schools for which there were up to five cohorts of available data. In all, there were 1,019 school cohorts and over 70,000 students with ACT Explore, ACT Plan, and the ACT test scores from three time points (Grades 8, 10, and 11 /12, respectively). For more details on the data, see the full report (Allen et al., 2009).

Method. Two general methods were used to estimate the effect of schools on ACT scores. The first method estimates the effect of schools on ACT scores, explicitly controlling for ACT Explore scores as covariates in a regression model, which do not require vertically-scaled assessments. The second method requires ACT Explore, ACT Plan, and ACT scores (referred to here as "EPA" scores) to be vertically-scaled and estimates the effect of schools on growth trajectories; that is, the degree to which attending a particular school affects students' score trajectories from Grade 8 to Grade 10 to Grades 11/12. For each method, two approaches were examined: one estimated school effects irrespective of contextual factors (referred to here as "ACT-VAM" and "EPA-VAM"). The second estimated

[2] This table is presented in ACT (2007b).

school effects, after making context adjustments for student-level factors such as family income and race/ethnicity, and school-contextual factors (referred to here as “ACT-CAVAM” and “EPA-CAVAM”). Estimates of school effects are referred to as “value-added” measures. School-contextual factors included Grade 11 enrollment, proportion of students tested, school poverty level, proportion of racial/ethnic minority students, and mean number of ACT Explore Benchmarks met. For more details on the methods, see the original report (Allen et al., 2009).

Results. Table 13.4 summarizes the distributions of the value-added measures generated by ACT-VAM and ACT-CAVAM, respectively. Distributions of these measures for the two models are very similar, but there is slightly less variation in the context-adjusted measures. Because the contextual factors (student-level and school-level) explain some of the variation in ACT scores across high school cohorts, there is less to be attributed to the school itself. Hence, the standard deviations of the context-adjusted value-added measures are smaller than the corresponding standard deviations of the unadjusted value-added measures.

Distributions of the EPA-VAM and EPA-CAVAM measures are reported in Table 13.5, indicating the similarity between the distributions of the two methods, with slightly less variation in the context-adjusted measures than the corresponding unadjusted effects. For both methods, the context-adjusted effects are highly correlated with the unadjusted effects, suggesting that value-added measures are less influenced by contextual factors.

Measures of the *uncertainty* of the value-added measures were established using *p*-values. Each high school cohort was classified as below average (estimated effect < 0, *p*-value < .05), above average (estimated effect > 0, *p*-value < .05), or uncertain (*p*-value > .05). Most of the estimated school effects from the four models could not be classified as “below average” or “above average” with reasonable certainty. For example, for value-added measures generated from the ACT-VAM model, 66% (for English) to 83% (for science) are classified as uncertain. For the ACT-CAVAM model, 67% (for English) to 86% (for science) of the school effects are classified as uncertain. Similar results are obtained for the EPA-VAM and EPA-CAVAM models. This shows that most school effects are not significantly different from the “average” school effect and cannot usually be distinguished from “average” with certainty.

For the value-added measures generated by the EPA-VAM and EPA-CAVAM models, the relationships with prior mean academic achievement and school characteristics are very similar to those observed for the ACT-VAM and ACT-CAVAM models. Surprisingly, prior mean academic achievement level (mean ACT Explore Benchmarks met) is negatively related to the value-added measures. Thus, cohorts with higher entering student achievement levels had significantly lower value-added scores.

Summary. The results of the analyses show 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. Thus, the most common scenario for a high-stakes decision based on value-added measures is that no action (rewarding or sanctioning) should be taken. This study highlights the need for reporting the statistical uncertainty about estimates of schools’ effects so that results can be properly interpreted.

Table 13.4. Distributions of Estimated School Effects on ACT Scores

	Estimate of school effect on ACT score					
Subject	Min	P_{25}	Med	P_{75}	Max	SD
			ACT-VAM method			
English	-2.62	-0.61	-0.02	0.61	2.74	0.91
Mathematics	-2.47	-0.51	-0.01	0.50	2.22	0.75
Reading	-1.79	-0.40	0.01	0.37	2.24	0.59
Science	-1.57	-0.33	0.00	0.31	1.74	0.48
			ACT-CAVAM method			
English	-2.59	-0.60	-0.03	0.57	2.76	0.86
Mathematics	-2.20	-0.48	-0.01	0.44	2.32	0.69
Reading	-1.79	-0.36	0.03	0.34	1.74	0.55
Science	-1.44	-0.28	0.01	0.28	1.34	0.43

Table 13.5. Distributions of Estimated School Effects on EPA Growth Trajectories

	Estimate of school effect on EPA growth trajectories					
Subject	Min	P_{25}	Med	P_{75}	Max	SD
			EPA-VAM			
English	-0.67	-0.15	0.00	0.15	0.66	0.23
Mathematics	-0.65	-0.14	-0.01	0.13	0.64	0.20
Reading	-0.47	-0.10	0.00	0.09	0.55	0.15
Science	-0.43	-0.09	0.00	0.08	0.42	0.13
			EPA-CAVAM			
English	-0.67	-0.15	-0.01	0.14	0.69	0.22
Mathematics	-0.58	-0.13	0.00	0.12	0.58	0.18
Reading	-0.45	-0.09	0.00	0.09	0.43	0.14
Science	-0.37	-0.07	0.00	0.07	0.34	0.11

13.5.2.2 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.

Data. Six data sets, corresponding to six different growth periods, were created from longitudinal test score data from the academic years 2006–2007 through 2011–2012. The growth periods included: Grades 8–10, 9–10, 10–11, 10–11/12 (10–12 for brevity), and 8–11/12 (8–12 for brevity). Note that Data Set 2 (Grade 9–10, 10–14-month interval) is a subset of Data Set 3 (Grade 9–10, 9–18-month interval). The ACT assessment system included ACT Explore (for Grades 8 and 9), ACT Plan (for Grade 10), and the ACT test (for Grades 11 and 12). Sample sizes and demographic breakdowns of each of the six data sets are presented in the original report (Bassiri, 2015).

Method. Three types of conditional status models were evaluated based on the accuracy with which they predicted growth across six different growth periods. The models included two hierarchical linear regression models (with and without adjusting for covariates) and a quantile regression model. Because prediction accuracy varied very little across the three methods, the hierarchical linear regression model without additional covariates was used to address the research questions set forth in this study. For discussion on the other models, please see the original report (Bassiri, 2015).

Test 2 (ACT Plan or the ACT test) scores were regressed on student's test 1 (ACT Explore or ACT Plan) scores in all four subject areas, the number of months that passed between the two tests, and the prior subject-specific mean ACT Explore (Plan) score for the high school using a two-level hierarchical linear regression model (Raudenbush & Bryk, 2002) with random intercepts. The value-added score for a particular school was calculated as the average of the residual scores of its students. The one-sample t-test can be used to test whether the mean residual is different from 0 for each school. Multiple linear regression was also used to assess the relationships of school characteristics and value-added scores.

Result. As expected, the projection parameters for the four prior test scores tend to be higher for the same-subject regression coefficients. There are some exceptions: for example, for predicting ACT Plan science scores, the estimated regression coefficients for Grades 9–10 are comparable for the ACT Explore mathematics score (0.309) and the ACT Explore science score (0.311). The estimated regression coefficients for the testing span are all positive (ranging from 0.029 to 0.120), indicating that more growth is expected as more time passes between the two testing periods.

Table 13.6 summarizes the distributions of the value-added measures (i.e., mean residual scores) for each subject area and growth period. The largest variation in value-added scores was obtained for Grades 8–12 (*SD* ranging from 0.99 to 1.18) and Grades 10–12 (*SD* ranging from 0.78 to 0.91). The number of students needed for scores at the 75th and 90th percentiles to be statistically significant was calculated. The value-added score estimates (at the 75th and 90th percentiles) obtained for Grades 8–12 are higher than those obtained from the other five data sets and require relatively fewer students in order for the estimates to be statistically significant. The results suggest that a small school (e.g., with 50 tested students) would need to have a value-added score above the 75th percentile in order to reach statistical significance (Table 13.6).

Table 13.6. School Cohort Value-Added Score Distributions

Data Set	Growth period	Range of month span	Subject	SD	P_{25}	P_{50}	P_{75}	P_{90}	N_{75}	N_{90}
1	8–10	18–30	English	0.64	-0.37	0.03	0.43	0.81	134	37
			Mathematics	0.81	-0.58	-0.09	0.46	1.02	140	28
			Reading	0.76	-0.48	-0.01	0.47	0.94	168	41
			Science	0.65	-0.43	-0.02	0.40	0.80	142	35
2	9–10	10–14	English	0.65	-0.31	0.06	0.42	0.80	129	36
			Mathematics	0.77	-0.54	-0.09	0.39	0.83	197	44
			Reading	0.78	-0.40	-0.01	0.38	0.75	251	65
			Science	0.64	-0.36	0.02	0.38	0.71	154	46
3	9–10	9–18	English	0.59	-0.31	0.06	0.42	0.78	128	35
			Mathematics	0.70	-0.53	-0.08	0.39	0.82	197	43
			Reading	0.64	-0.39	0.00	0.37	0.73	233	62
			Science	0.56	-0.34	0.03	0.38	0.70	152	44
4	8–11/12	30–54	English	1.14	-0.70	0.01	0.76	1.48	78	20
			Mathematics	1.18	-0.71	-0.01	0.83	1.70	50	12
			Reading	1.00	-0.64	0.06	0.70	1.25	109	35
			Science	0.99	-0.59	0.06	0.69	1.27	83	24
5	10–11	10–14	English	0.89	-0.60	0.04	0.56	1.07	110	30
			Mathematics	0.73	-0.50	-0.07	0.43	0.93	127	28
			Reading	0.73	-0.45	0.02	0.48	0.89	198	56
			Science	0.70	-0.47	0.01	0.48	0.82	144	48
6	10–11/12	9–30	English	0.91	-0.57	0.02	0.62	1.17	98	27
			Mathematics	0.88	-0.54	0.00	0.63	1.23	68	18
			Reading	0.81	-0.52	0.01	0.53	0.98	170	49
			Science	0.78	-0.47	0.06	0.56	0.98	108	35

Across all growth periods, projection accuracies were highest for English and lowest for science. However, the proportion of projected scores that were within 1, 2, or 3 points of the actual test 2 score was higher for science than for English. This is due to the fact that science has the smallest standard deviation of the four subject areas; therefore, the absolute distance between projected and actual scores for science tends to be smaller.

Table 13.7 presents the correlation (R) between projected and actual ACT scores, and the proportion of projection accuracy (W_1, W_2, and W_3)[3] of two different models predicting ACT scores. In the first model, the four ACT Plan test scores were used as predictor variables, whereas in the second model, the four ACT Explore scores and the four ACT Plan scores were used as predictors. This reveals that projection accuracy is slightly enhanced in the second model.

Table 13.7. Comparison of Projection Accuracy (ACT Plan vs. ACT Explore and ACT Plan Prior Test Scores)

Growth period	Range of month span	Subject	R	W_1	W_2	W_3
10–11/12 (with Plan)	9–30	English	0.85	0.25	0.48	0.66
		Mathematics	0.84	0.27	0.52	0.71
		Reading	0.81	0.23	0.43	0.61
		Science	0.79	0.27	0.51	0.70
10–11/12 (with Plan and Explore)	9–30	English	0.87	0.27	0.50	0.69
		Mathematics	0.86	0.29	0.54	0.73
		Reading	0.83	0.24	0.45	0.63
		Science	0.80	0.28	0.53	0.71
n = 525,194 students						

Note. R is the correlation between projected and actual ACT scores.
W_1 is the proportion of ACT projected scores that are within 1 point of the actual score.
W_2 is the proportion of ACT projected scores that are within 2 points of the actual score.
W_3 is the proportion of ACT projected scores that are within 3 points of the actual score.

Results from measures of the uncertainty of the estimated school effects (mean residual scores) corroborated the results obtained from an earlier study (Allen et al., 2009). It suggests that most estimated school effects should not be classified as “below average” or “above average” with high confidence. Most school effects are not significantly different from the average school effect and cannot usually be distinguished from average with a high degree of confidence.

Simple correlations between value-added measures (i.e., mean residual scores) and school characteristics shows that prior mean academic achievement is positively related to the value-added measures ($r = 0.13$ to $r = 0.52$), whereas school poverty level and proportion of racial/ethnic minority students have inverse relationships with the school effects ($r = -0.29$ to $r = -0.56$ and $r = -0.08$ to $r = -0.45$, respectively).

[3] W_1 is the proportion of ACT projected scores that are within 1 point of the actual score. W_2 is the proportion of ACT projected scores that are within 2 points of the actual score. W_3 is the proportion of ACT projected scores that are within 3 points of the actual score.

Generally, prior mean academic achievement level was positively related to the value-added measures, whereas poverty level was inversely related to the value-added measures. Thus, cohorts with higher entering student achievement levels, as well as cohorts from wealthier schools, had significantly higher value-added scores. For the growth period ending in Grade 10, prior mean academic achievement level has a larger positive effect (ranging from $b = 0.24$ to $b = 0.40$) in predicting growth than for growth periods ending in Grade 11 or 12. This may be due to the use of ACT Plan scores (Grade 10) instead of the ACT scores (Grade 11 or 12). For growth periods ending in Grade 11 or 12, school poverty level had a larger negative effect (ranging from $b = -0.54$ to $b = -0.24$). Generally, compared to other school characteristics, class size and proportion of students tested had weaker associations with value-added measures.

Cross-cohort correlations of the value-added measures for adjacent cohorts (one year apart), as well as for cohorts that are two and three years apart reveals that the Grades 8–12 value-added measures have substantially greater consistency over time, relative to those measured over shorter periods.

Summary. Value-added scores based on longer timeframes (i.e., Grades 8–12) are more likely to distinguish school effects. The results also underscore the influence of prior academic achievement, particularly in the same subject but including off-subject scores, on future scores. Of school characteristics, prior mean academic achievement is positively related to the value-added measures, whereas school poverty level and proportion of racial/ethnic minority students have 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 is the outcome variable, poverty level and class size tend to be more predictive of value-added scores. When ACT Plan is the outcome variable, prior mean academic achievement tends to be more predictive. Value-added scores for low-poverty schools were higher than those obtained from high-poverty schools in all subject areas.

13.5.2.3 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 to year two.

Sample. The sample comprised 1,119 high schools and 263,737 students, who had test scores from two time points (ACT Explore in eighth grade and the ACT in eleventh or twelfth grade). For each high school, there was up to six cohorts of data available, representing the graduating classes of 2004 to 2009. In all, there were 2,707 cohort-by-high school combinations. The student sample was quite typical of ACT Explore-tested or the ACT-tested populations in terms of academic achievement. The sample of high schools was similar to the population of public high schools with respect to poverty level but had relatively fewer high-minority schools. Student demographics are presented in the original paper.

College Enrollment and Retention Data. Data from the National Student Clearinghouse (NSC) were used to identify students who enrolled in college the fall after high school graduation (first-year enrollment) and who re-enrolled at the same or a different postsecondary institution the second fall after high school graduation (retention).

College Course Grade Data. First-year college course grade data were collected across multiple years from postsecondary institutions participating in ACT's Course Placement or Prediction Services. In all, there were 26,863 students with first-year college course data. In the study sample, only courses from four core content areas (English/language arts, mathematics, social sciences, and natural sciences; 83% of all available course data) were considered. The breakdown of the four core content areas was 26% from English/language arts, 17% from mathematics, 14% from natural sciences, and about 26% from social sciences. The remaining 17% of course data that was coded as noncore included 6% from fine arts, 1% from business, about 1% from foreign languages, and 9% from the miscellaneous category. By course type, the highest enrollment was in Composition I (15%), followed by Composition II, College Algebra, and American History (7% each); the lowest enrollments were in Archaeology and Geometry.

Method. High schools' effects on ACT scores were estimated using a hierarchical linear regression model. The school effect can be interpreted as the number of ACT score points attributable to a school, above and beyond what can be attributed for the average school. The model controlled for prior academic achievement level as measured by the same students' ACT Explore scores in eighth grade in four areas, the number of months between ACT Explore and the ACT testing, gender, race/ethnicity, and school characteristics (school size, proportion of students tested, poverty level, proportion of racial/ethnic minority students, and mean ACT Explore scores).

A two-level hierarchical logistic regression model with random intercepts was used to predict college enrollment and college retention (binary outcomes), and a two-level hierarchical linear regression model (Raudenbush & Bryk, 2002) with random intercepts was used to predict college course grades. In the college course grades model, the grades in first-year college courses were regressed on the school effect estimates and student- and school-level covariates. Separate models were fit for each of the four core college content areas (English/language arts, mathematics, natural sciences, and social sciences). For college enrollment, the models treated students as nested within high schools. For the college outcomes (retention and grades), the models treated students as nested within colleges.

Enrollment and Retention Results. The point-biserial correlations between college enrollment and retention and characteristics of schools and students were all statistically significant at $p < .0001$. The high school effect measures are positively related to college enrollment and retention. The correlation coefficients also indicate that students' eighth grade academic achievement has the strongest relationships with enrollment and retention among the variables studied. The correlations are larger for enrollment (0.19 to 0.21) than for re-enrollment (ranging from 0.14 to 0.16), suggesting that retention is less influenced by students' prior academic achievement.

The high school effect measure was associated with higher log-odds of enrollment. For each one standard deviation increase in the high school effect measure, the log-odds increase by 0.14, indicating that the odds of enrollment increase by a factor of 1.15 ($e^{0.14}$) for each standard-deviation increase in the high school effect measure.

The baseline predicted probabilities of re-enrolling at any or at the same institution are 0.88 and 0.73, respectively. The retention probabilities increase with the high school effect estimate, school size, proportion tested, and proportion minority; and is greater for those with higher ACT Explore scores. The model suggests that high school effect measure is associated with a higher log-odds of retention at any or at the same institution, by 0.07 and 0.03, respectively. ACT Explore scores, especially in mathematics, were positively and significantly related to enrollment (log-odds = 0.23) and re-enrollment at any (log-odds = 0.18) and the same college (log-odds = 0.09).

College Course Grades Results. Based on bivariate correlational results, the high school effect estimates were not significantly correlated with college course grades in English/language arts, mathematics, and natural sciences and were only weakly correlated with grades in social sciences (discussed more below). On the other hand, ACT Explore scores had the strongest correlations with grades in each content area, ranging from 0.19 to 0.25. At the school level, the mean ACT Explore scores were correlated with college course grades in each respective subject area, with correlations of 0.09 or 0.10. High school poverty level, proportion minority, and time between ACT Explore and the ACT testing were negatively related to college course grades across all content areas.

All regression coefficient estimates of the high school effects were positive and statistically significant; ranging from 0.02 in English/language arts (significant at $p < .05$); 0.04 in natural sciences and social sciences; to 0.07 in mathematics (all significant at $p < .01$). So while correlations of the high school effects and grades were not significant, the high school effects are significant predictors of grades in a model that controls for other covariates.

Summary. The study found that value-added measures representing school effects on ACT scores have small but 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 also found that the majority of the variance in college enrollment, retention, and in first-year course grades is due to students' characteristics; and less of the variance is due to the characteristics of high schools or colleges. This was evidenced by their statistically significant but small intraclass correlation coefficient (ICC) estimates.

13.6 ACT Test-Retest Statistics

13.6.1 Retesting with the ACT

Increasing numbers of students are taking 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 score gains with the ACT Composite score and focused on how likely students are to obtain or exceed a specific ACT Composite score on retesting given their initial score. In this investigation, the mean gain on retesting was found to be 0.8 scale score points. A follow-up study (Andrews & Ziomek, 1998) 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. Approximately 95% of all students had a 70% to 80% chance of maintaining or increasing their score on retesting. The percentage of examinees maintaining or increasing their score, as well as the amount of the average gain, decreased with each additional testing. The average ACT Composite score gain on retesting was 0.75 points. As illustrated in Figure 13.3, students with lower scores on previous tests had the greatest average gains and those scoring near the maximum score of 36 actually had score decreases. Figure 13.4 shows the percentage of students maintaining or increasing their scores over multiple tests.

More recently, 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.

Data and method. The sample included 1,924,436 students from the 2015 graduating high school class. Single test takers numbered 1,054,773; students who took the ACT test two times numbered 504,222; students who tested three times numbered 218,521; and students who tested four or more times numbered 146,920.

Results. Most students (78%) who retested improved or maintained their ACT Composite score on the second test. The average final ACT Composite score was consistently higher as the number of times students tested increased. As found by Andrews and Ziomek (1998), the percentage of students who increased their scores upon retesting was higher when their initial score was low, as compared to gains made by students whose initial scores were high.

An even more prominent factor associated with score gains was time between testing (Harmston & Crouse, 2016). As time between testing increases, the potential for greater curricular coverage to occur in the interval between tests increases. That is, students may have the opportunity to master more of the tested material in their classes. Using grade level as a proxy for curriculum coverage and additional time for test preparation, 2015 graduates who first tested as sophomores (N = 79,346) saw an average ACT Composite score increase of 2.7 points by their final test session. Students first testing as juniors (N = 695,502) demonstrated an average score increase of 1.1 points. Students taking their first and last tests as seniors (N = 93,695) gained only 0.6 points on average.

Summary. 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.

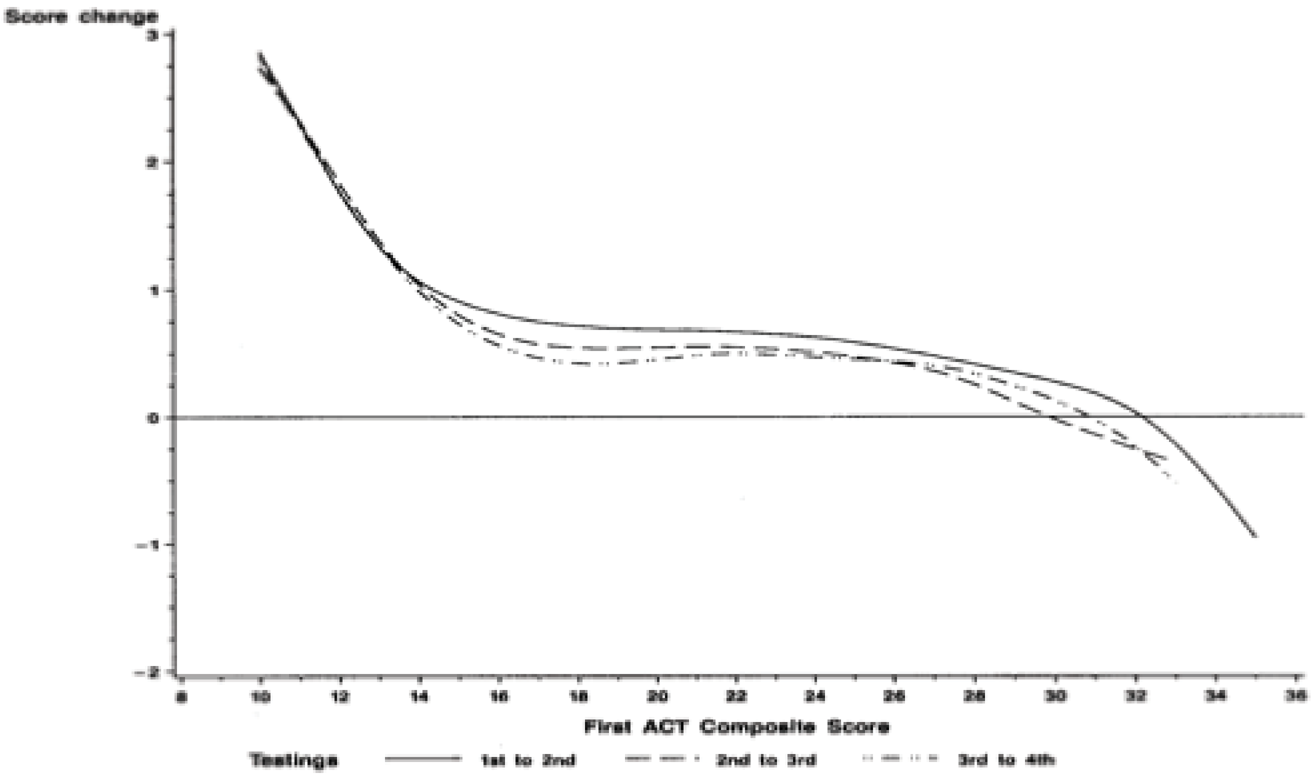


Figure 13.3 Changes in Composite test scores from 1st to 2nd, 2nd to 3rd, and 3rd to 4th testing

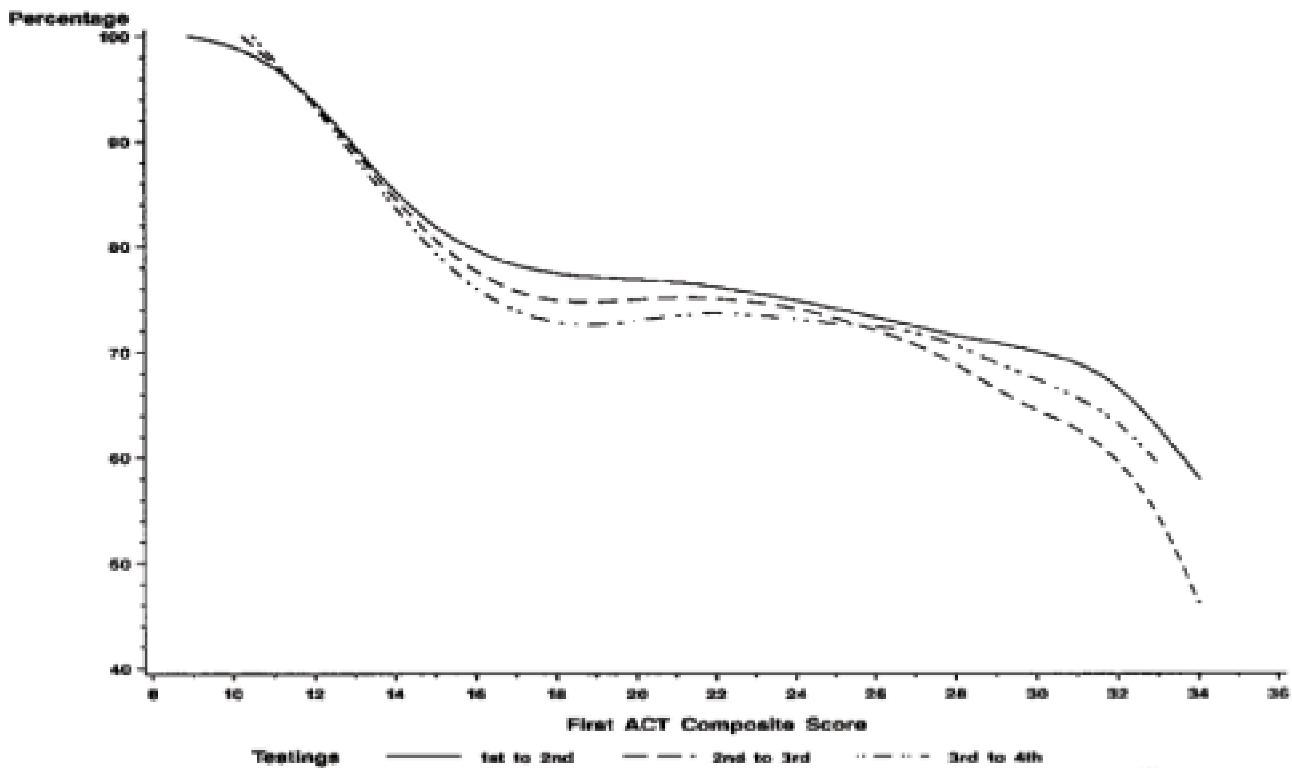


Figure 13.4 Percentage of students maintaining or increasing score from 1st to 2nd, 2nd to 3rd, and 3rd to 4th testing

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. Table 13.8 presents summary retest statistics by initial 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 is 1 point.

Table 13.8. ACT Composite Score Retest Statistics, by Initial ACT Composite Score

ACT Composite score from first test	ACT Composite score from second test		Percentage of students whose scores changed or remained the same from first to second test *		
	Typical score	Range for middle 50%	Increased	Remained the same	Decreased
35**	35	34 to 35	16	41	43
34**	34	33 to 35	33	32	35
33	33	32 to 34	41	27	31
32	32	31 to 33	46	24	30
31	31	30 to 32	48	24	28
30	30	29 to 32	50	23	27
29	30	28 to 31	51	23	26
28	29	27 to 30	53	21	25
27	28	27 to 29	54	21	24
26	27	26 to 28	55	22	24
25	26	25 to 27	55	22	23
24	25	24 to 26	56	22	22
23	24	23 to 25	56	22	22
22	23	22 to 24	57	21	22
21	22	21 to 23	57	21	22
20	21	20 to 22	57	21	22
19	20	19 to 21	57	20	22
18	19	18 to 20	58	20	22
17	18	17 to 19	57	20	23
16	17	16 to 18	58	20	22
15	16	15 to 17	59	20	21
14	15	14 to 16	61	20	19
13	14	13 to 15	67	20	14
12	14	13 to 15	76	17	7
11	13	12 to 14	88	9	4

* Percentages may not sum to 100 due to rounding.

** Results for these ACT Composite scores are based on a relatively small number of students.

References

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

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

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

Allen, J., Bassiri, D., & Noble, J. (2009). *Statistical properties of accountability measures based on ACT's educational planning and assessmen*t. (ACT Research Report No. 2009-1). Iowa City, IA: ACT.

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

Andrews, K. M., & Ziomek, R. L. (1998). *Score gains on retesting with the ACT Assessment.* (ACT Research Report No. 98-7). Iowa City, IA: ACT, Inc.

Bassiri, D., Allen, J. (2012). *Grade 8 to 12 Academic growth patterns for English language learners and students with disabilities*. (ACT Research Report No. 2012-1). Iowa City, IA: ACT, Inc.

Bassiri, D. (2014). *Research study: The forgotten middle*. (ACT Policy Brief 2014). Iowa City, IA: ACT, Inc.

Bassiri, D. (2015). *Statistical properties of school value-added scores based on assessments of college readiness*. (ACT Research Report No. 2015-5). Iowa City, IA: ACT, Inc.

Bassiri, D. (2016a). *Academic growth patterns of first-generation college students in grades 8 to 12.* (ACT Research Report No. 2016-6). Iowa City, IA: ACT, Inc.

Bassiri, D. (2016b). *Are value-added measures of high school performance related to students' enrollment and college success?* (ACT Research Report No. 2016-10). Iowa City, IA: ACT, Inc.

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.

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

Castellano, K. E., & Ho, A. D. (2013). A practitioner's guide to growth models. Washington, DC: CCSSO. Retrieved March 8, 2017 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.

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

Holland, J. L. (1997). *Making vocational choices: A theory of vocational personalities and work environments*. (3rd ed.). Odessa, FL: Psychological Assessment Resources.

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

Lanier, C. W. (1994). *ACT composite scores of retested students*. (ACT Research Report No. 94-3). Iowa City, IA: American College Testing Program.

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

Sawyer, R. L. (2008). *Benefits of additional high school course work and improved course performance in preparing students for college.* (ACT Research Report No. 2008-1). Iowa City, IA: ACT, Inc.

Wai, J., & Allen, J. (2017). Examining predictors of academic growth in secondary school among academically advanced youth across 21 years. Unpublished manuscript.

CHAPTER 14

Other ACT Components

14.1 The ACT Interest Inventory

14.1.1 Overview

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 especially critical. Using their interest inventory results, students can explore programs of study and occupations in line with their activity preferences.

The ACT Interest Inventory consists of 72 items and provides scores on six scales paralleling 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, 2009). The current 72-item edition of the inventory is referred to in that manual as UNIACT-S.

14.1.2 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 Grade 12 nationally representative sample involving over 250,000 students from over 8,000 schools (for more information on the development of these norms, see ACT, 2009). These scores are made available for counselors who are familiar with Holland's theory of career types (Holland, 1997) and who want to use these scores to offer a clinical interpretation of the student's interests.

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 as well as the results of measured interests.

ACT Interest Inventory results are reported on the ACT Student Report in two ways. First, it includes a short list of occupations that primarily involve the kinds of basic work tasks that the student prefers. Second, it displays the results from the ACT Interest Inventory on the Career Connector (shown in Figure 15.1 of Chapter 15). The Career Connector is a two-dimensional figure with four compass points labeled Working with People, Data, Things, and Ideas (see ACT, 2009 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 arrows on a Career Connector may show that the student primarily enjoys activities involving ideas and people. The Career Connector is derived from the ACT's 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 actionable assessment results.

Career Map. The ACT Career Map (Figure 14.1) 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 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, 2009; Prediger & Swaney, 2004). Purpose of the work and work setting were also considered when the Career Areas were formed.

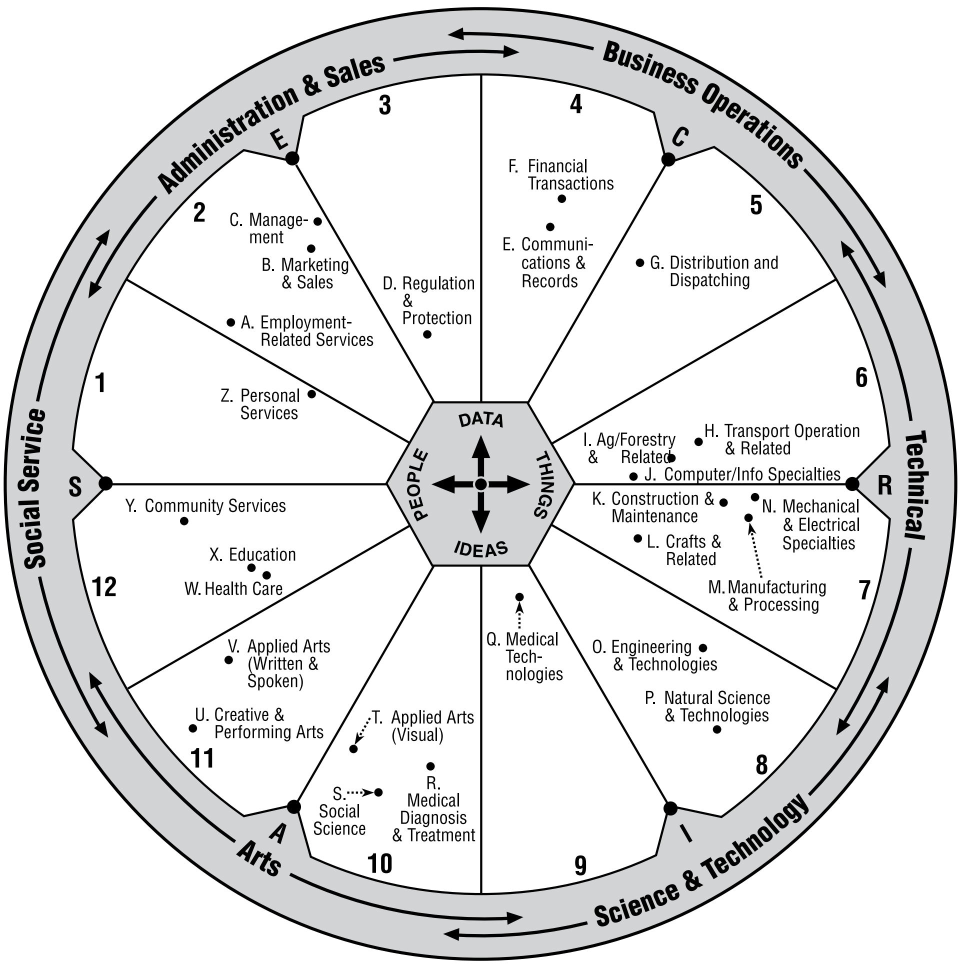


Figure 14.1 The ACT Career Map.

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 14.1, 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, 2009).

14.1.3 Psychometric Support

The ACT Interest Inventory Technical Manual (ACT, 2009), 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 (Mdn = .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).

14.1.4 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. Level of interest-major fit is displayed on the Student, High School, and College score reports as shading of one of the three (Low, Medium, or High) sections of the Interest-Major Fit Bar (see Figure 15.1 in Chapter 15).

Evidence clearly indicates that the fit between students' interests and their college majors is important in 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, 2013). 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, involving data over a 60-year time period and including 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 (2009).

14.2 The High School Course/Grade Information Section

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. Students registering for national test dates 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. High school GPAs based on self-reported grades are shown on the ACT College Report for English, mathematics, natural sciences, and social studies. 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 11.

14.3 The Student Profile Section

In addition to measures of educational development and high school grades, other student information is collected as part of registration for the ACT to broaden the information base of both students and colleges. The development of the Student Profile Section (SPS) has been influenced by the educational context in which it evolved, as have other parts of the ACT. The primary assumption underlying development of the SPS is that the quality of education provided depends, in part, on the amount of relevant information a college has about its students. The SPS is intended to make this information available in a systematic form prior to enrollment.

The SPS contains several subsections, each of which is discussed below. The items of the SPS have been developed by ACT staff with input from personnel from a variety of postsecondary educational institutions. Items are revised from time to time as needs arise for these institutions to obtain different types of data.

Admissions/Enrollment Information. The questions in this subsection of the SPS are designed to yield two types of information. The first type is essential to planning by colleges since it includes the student's enrollment plans (full-time/part-time and preferred type of living accommodations). The second type of information relates to the presence of any physical disability or learning disability. The instructions explicitly state that disability information need not be supplied.

Educational Plans, Interests, and Needs. 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 SPS provides opportunities for students to indicate such information 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 their students.

Educational Needs and Interests. With each new entering class, colleges must be prepared to provide individualized assistance to support the academic development of their students. In this subsection, students indicate their 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 are able to alert the college about their individual needs. This subsection also includes questions about student interest in college programs designed for enriched or accelerated academic work.

College Extracurricular Plans. 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 plans is another way of communicating their unique patterns of interests, needs, and skills. The information provided in this subsection of the SPS includes interest in social, political, and religious organizations, as well as the arts, athletics, and other activities.

Financial Aid. Questions about the student's plans for financing a college education make up this subsection of the SPS. Student responses to these questions can be useful to college financial aid officers. Students are asked to estimate the family's annual income and to indicate if they intend to apply for financial aid and/or to work part-time while in college.

Background Information. This subsection of the SPS focuses on religious affiliation, distance from college, language spoken in the home, racial/ethnic background, and parent/guardian education level. Because some individuals may prefer not to supply this information, several of these questions provide students with the option to not respond. The information collected from this subsection is intended to be used by colleges in the planning process.

Factors Influencing College Choice. Information about how the student chooses a college can be of use to personnel responsible for planning. This subsection of the SPS contains questions about the type of institution (public/private, coeducational or not, and two-year/four-year), size, location, and maximum tuition that the student prefers in a college. The student is also asked to rank these factors, along with the college curriculum, in order of importance.

High School or Home School Information. This subsection of the SPS asks the student to supply information about the type of high school attended (public/private). If the student will graduate from a home school, this subsection asks how many years of high school homeschooling the student will have completed by graduation.

High School Information. This subsection requests information about the student's high school performance (overall GPA, class rank) and the number of students in her or his graduating class. 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). Finally, students are asked to indicate high school activities they have participated in. Students select from a list of activities such as athletics, drama, music, student government, student publications, and special-interest clubs.

Out-of-Class Accomplishments. Accomplishments (e.g., awards, election to offices, creative productions) in extracurricular activities while in high school are the focus of this subsection of the SPS. In conjunction with the questions in the previous subsection, these questions allow the student to report particular achievements as well as participation in a wide range of out-of-class activities.

References

ACT. (2009). *The ACT interest inventory technical manual*. Retrieved from http://www.act.org/content/dam/act/unsecured/documents/ACTInterestInventoryTechnicalManual.pdf

ACT. (2013). *College choice report part 3: Persistence and transfer*. Iowa City, IA: Author.

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*, 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,* 23–35.

Holland, J. (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* (ACT Research Report No. 29). Iowa City, IA: ACT, Inc.

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*, 384–403.

Prediger, D. (1996). Alternative dimensions for the Tracey-Rounds interest sphere. *Journal of Vocational Behavior, 48*, 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*, 440–459.

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

CHAPTER 15

Self-Report Information on ACT Score Reports

15.1 Overview

This chapter briefly describes selected sections on the student, high school, and college score reports. The focus of this chapter is primarily on student self-report information, but also includes some results that integrate self-report information with information derived from test scores. All of these different types of information share a common purpose in that they assist students, directly or indirectly, in charting a path to college and career. Test scores and related indicators, also relevant to this purpose, are described in Chapter 7.

The enhanced score reports, introduced in the fall of 2016, allow students and educators to better navigate results and gain more meaningful insights. ACT has provided training and new support materials to assist schools and districts in using these reports. The ACT Test User Handbook for Educators provides descriptions of all sections included on the enhanced ACT score reports. The User Handbook can be found at: http://www.act.org/content/act/en/products-and-services/the-act-educator/resources.html.

15.2 Student Report

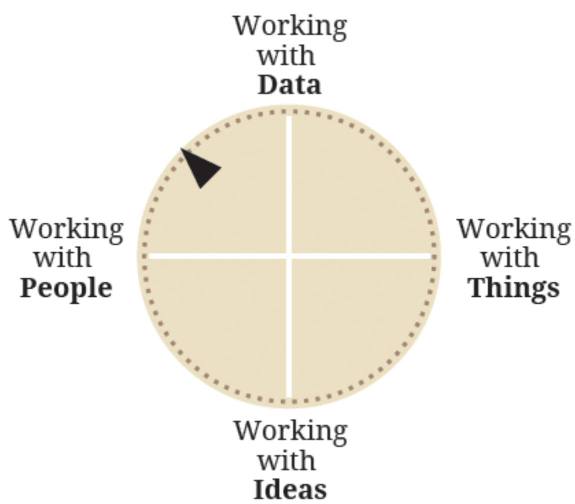
College and Career Planning. This section of the student score report provides a visual summary of the results of the ACT Interest Inventory. The Career Connector is a two-dimensional figure with four compass points, labeled Working with People, Data, Things, and Ideas. These are the four basic work tasks shown to underlie the work done in all occupations (ACT, 2009). The Career Connector summarizes the pattern of results from the scales on the ACT Interest Inventory and visually displays it as one or two directions with respect to these compass points. As shown in Figure 15.1, the Career Connector also includes a personalized list of five occupations that involve work tasks in this same direction.

College and Career Planning

Where are you going? Knowing your interests can help you find the kinds of majors and occupations that may be right for you. Occupations differ widely in how much they involve working with four basic work tasks: Data, Ideas, People, and Things. Before you took the ACT, you completed an interest inventory. Your results point to occupations that involve the kinds of basic work tasks you prefer. Visit **www.actprofile.org** to learn more.

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

- Buyer
- FBI/CIA Agent
- Financial Manager
- Training/Education Manager
- Travel Guide



Interest–Major Fit

Do your interests fit the college major you plan to enter? Based on information you provided, you plan to enter **Accounting**.

Your interests are fairly similar to the interests of college students in the major you plan to enter. Students in majors that fit their interests are often more satisfied with their major.

Figure 15.1 Two sections of the ACT Student Report.

The Career Connector is derived from the ACT's 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. The Career Map serves as an interpretive bridge linking people to occupations and is designed to engage users in the process of career exploration. Both the Career Map and the Career Connector are described in Chapter 14.

The Career Connector serves two roles. First, it visually displays a summary of the basic work-relevant interests of the student and lists several occupations that align with those interests. Second, it provides an introduction to the concepts that serve as the foundation for the ACT's Career Map. Thus the Career Connector, while not a map itself, summarizes ACT Interest Inventory results and helps students understand and use the Career Map to explore personally relevant career options.

Interest-Major Fit. As seen in Figure 15.1, the ACT Student Report lists the student's planned college major and displays the level of interest-major fit. The latter is shown by shading in one of the three sections (Low, Medium, or High) of the Interest-Major Fit Bar. The fit level is derived from two data elements: the student's ACT Interest Inventory scores and the self-reported major the student plans to enter. These elements are used to calculate an interest-major fit score, which is converted to one of three fit levels. 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. Chapter 14 describes interest-major fit in more detail and summarizes validity studies showing that interest-major fit predicts important student outcomes.

15.3 High School Report

The following are self-report sections on the High School Report that can be used to assist students in charting a path to college and career. In addition, both the High School Report and the Student Report include interest-major fit. Refer to Section 15.2 of this chapter for a description of interest-major fit.

ANN C TAYLOR
GRADUATION YEAR: 2018
TEST DATE: APRIL 2017 | NATIONAL

The ACT®

High School Report

College Fit

At the student's direction, scores from this test date are being reported to the colleges shown below. College planning information is provided for the choices listed when registered or tested. Check with colleges for recent changes in information. Note: GPA was calculated from the grades the student reported. For more information, see the *ACT User Handbook* at **www.act.org/the-act.html**.

College Choices	Chance of "B" or Higher GPA based on Composite score and GPA	ACT Composite Scores of the middle 50% of current students	Average High School GPA of current students	Average High School Class Rank of current students	Percentage of First-Year Students Receiving Financial Aid: Need	Merit
9521 UNIVERSITY OF OMEGA OMEGA, CO	61%	18–24 (scale 1–36)	3.26	Majority in Top 50%	67%	20%
9059 ALPHA UNIVERSITY UNIVERSITY CENTER, IA	43%	21–26 (scale 1–36)	3.38	Majority in Top 50%	85%	27%
8866 BETA COMMUNITY COLL CLARKSTON, CO	72%	16–21 (scale 1–36)	2.85	Majority in Top 50%	58%	18%
8905 MAGNA COLLEGE MAGNA, OH	56%	21–24 (scale 1–36)	3.23	Majority in Top 50%	90%	35%
		1–36				
		1–36				
ANN C TAYLOR		21 Composite	3.29 GPA	Top 25% Class Rank		

Educational and Occupational Plans

Interest Inventory Scores

Interest Area	Standard Score (20–80)
Science & Tech.	56
Arts	41
Social Service	59
Admin. & Sales	65
Business Oper.	51
Technical	42

Major*	Occupation*
Accounting	Economics
Fairly Sure	Very Sure

Interest–Major Fit

Low / **Medium** / High

To learn more, visit **www.actprofile.org**.

Examples of Related Majors	Examples of Related Occupations
Banking & Financial Support Services	Anthropologist
Business Administration/Mgmt, General	Criminologist
Finance	Gerontologist
Financial Planning & Services	Historian
Insurance	Political Scientist
Investments & Securities	Psychologist, Experimental
Purchasing/Procurement/Contract Mgmt	Sociologist
Entrepreneurship	Urban Planner

Needs Help With*

Educational/Occupational Plans, Writing, Reading, Study Skills, Math

Remember that **www.actprofile.org** can help students focus on careers, majors, and colleges that are right for them. It's **free**!

* Information provided by the student. If major and occupation boxes are not shaded, data was incomplete.

Figure 15.2 Side 2 of the ACT High School Report.

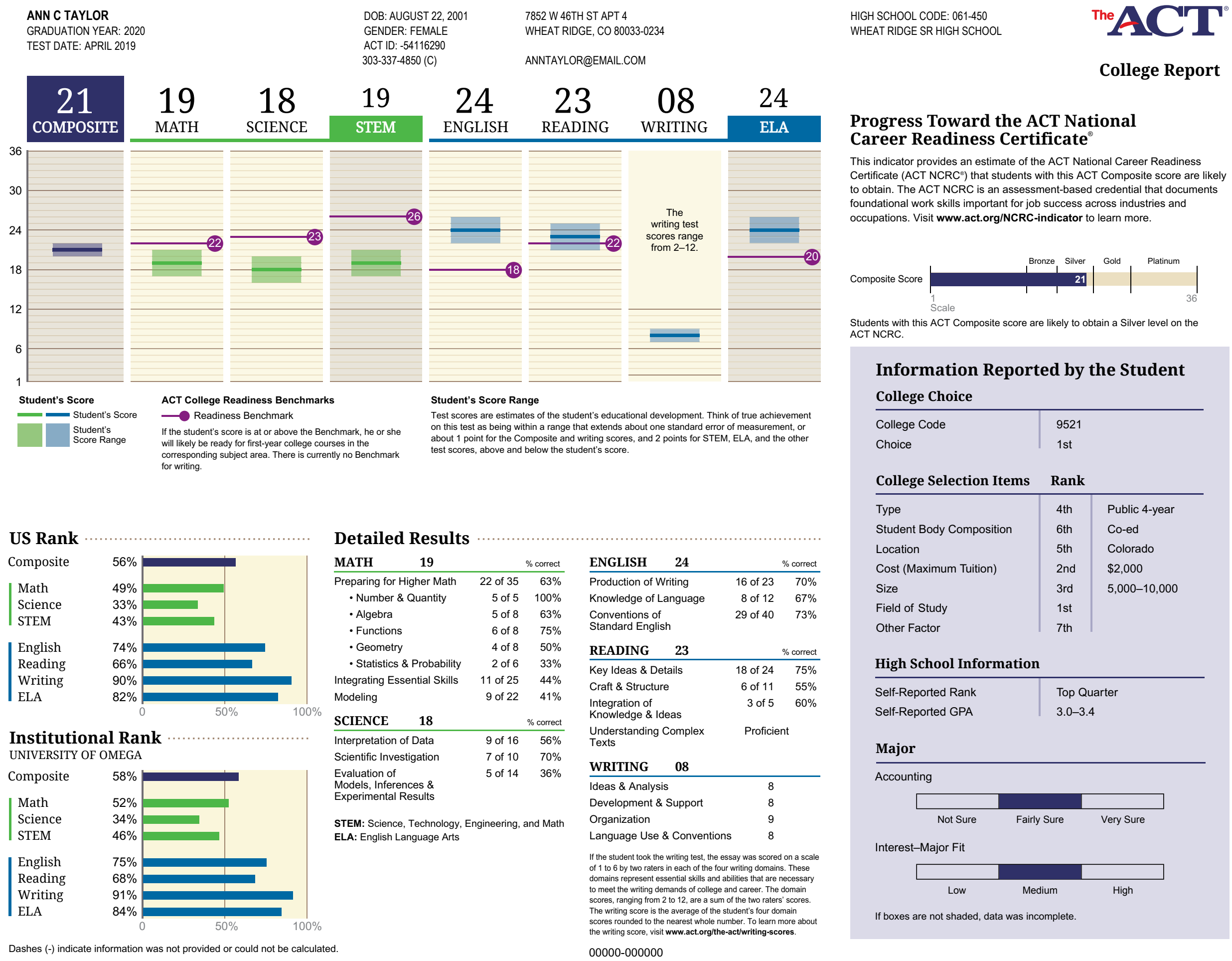
College Fit. The student's college choices are reported in this section along with information that can assist students in evaluating their fit with each of these colleges. The student's chance of attaining a GPA of 3.0 or higher, based on the student's ACT Composite score and self-reported GPA, is reported for each college. (Dashes indicate that the chances of success could not be calculated or that the college did not participate in the ACT Admissions Service.) Next, the student's ACT Composite score is compared to the Composite score range of the middle 50% of students at each college listed. Finally, average high school GPA and high school class rank of the current students for each college are noted so that the student can compare themselves to currently enrolled students at each college.

Educational and Occupational Plans. This section of the ACT High School Report consists of several parts (see Figure 15.2), each based—directly or indirectly—on self-reported information collected from students during ACT registration.

- ***ACT Interest Inventory Scores.*** The ACT Interest Inventory scores are reported as standard scores with a mean of 50 and a standard deviation of 10. Reliability and validity of the ACT Interest Inventory scores for career exploration and planning are well established (ACT, 2009). These scores are made available for counselors who are familiar with Holland's theory of career types (Holland, 1997) and may want to use these scores to offer a clinical interpretation of the student's interests.
- ***Major and Occupation Choices.*** The ACT High School Report lists the student's planned college major and choice of occupation. Also shown is the student's self-reported level of certainty (Not Sure, Fairly Sure, or Very Sure) for both their choice of major and occupation. Level of certainty has been found to be highly related to persistence of choice from planned major in high school to declared major in college (ACT, 2013).
- ***Examples of Related Majors and Occupations.*** This section lists up to eight majors related to the student's major they plan to enter and up to eight occupations related to their occupational choice. Many students consider several possibilities before making definite career plans. These lists can be used as a starting point for considering other possibilities to explore.
- ***Needs Help With.*** This is a list of educational areas the student reported needing help with. Students respond yes or no to each of the five areas listed in the Student Profile Section of the ACT: educational and occupational planning, writing, reading, study skills, and mathematics.

15.4 College Report

The College Report includes several elements found in the Student Report and/or the High School Report, including interest-major fit, major choice, and educational areas that the student reports needing help with. Refer to previous sections of this chapter for descriptions. The following are other sections on the College Report that can be used to assist students in charting a path to college and career.



ANN C TAYLOR
GRADUATION YEAR: 2020
TEST DATE: APRIL 2019

DOB: AUGUST 22, 2001
GENDER: FEMALE
ACT ID: -54116290
303-337-4850 (C)

7852 W 46TH ST APT 4
WHEAT RIDGE, CO 80033-0234

ANNTAYLOR@EMAIL.COM

HIGH SCHOOL CODE: 061-450
WHEAT RIDGE SR HIGH SCHOOL

The ACT®

College Report

COMPOSITE	MATH	SCIENCE	STEM	ENGLISH	READING	WRITING	ELA
21	19	18	19	24	23	08	24

The writing test scores range from 2–12.

Student's Score
Student's Score
Student's Score Range

ACT College Readiness Benchmarks
Readiness Benchmark
If the student's score is at or above the Benchmark, he or she will likely be ready for first-year college courses in the corresponding subject area. There is currently no Benchmark for writing.

Student's Score Range
Test scores are estimates of the student's educational development. Think of true achievement on this test as being within a range that extends about one standard error of measurement, or about 1 point for the Composite and writing scores, and 2 points for STEM, ELA, and the other test scores, above and below the student's score.

Progress Toward the ACT National Career Readiness Certificate®

This indicator provides an estimate of the ACT National Career Readiness Certificate (ACT NCRC®) that students with this ACT Composite score are likely to obtain. The ACT NCRC is an assessment-based credential that documents foundational work skills important for job success across industries and occupations. Visit **www.act.org/NCRC-indicator** to learn more.

Composite Score: 21 (Bronze, Silver, Gold, Platinum; Scale 1–36)

Students with this ACT Composite score are likely to obtain a Silver level on the ACT NCRC.

US Rank

Composite	56%
Math	49%
Science	33%
STEM	43%
English	74%
Reading	66%
Writing	90%
ELA	82%

Institutional Rank
UNIVERSITY OF OMEGA

Composite	58%
Math	52%
Science	34%
STEM	46%
English	75%
Reading	68%
Writing	91%
ELA	84%

Dashes (-) indicate information was not provided or could not be calculated.

Detailed Results

MATH 19		% correct
Preparing for Higher Math	22 of 35	63%
• Number & Quantity	5 of 5	100%
• Algebra	5 of 8	63%
• Functions	6 of 8	75%
• Geometry	4 of 8	50%
• Statistics & Probability	2 of 6	33%
Integrating Essential Skills	11 of 25	44%
Modeling	9 of 22	41%

SCIENCE 18		% correct
Interpretation of Data	9 of 16	56%
Scientific Investigation	7 of 10	70%
Evaluation of Models, Inferences & Experimental Results	5 of 14	36%

STEM: Science, Technology, Engineering, and Math
ELA: English Language Arts

ENGLISH 24		% correct
Production of Writing	16 of 23	70%
Knowledge of Language	8 of 12	67%
Conventions of Standard English	29 of 40	73%

READING 23		% correct
Key Ideas & Details	18 of 24	75%
Craft & Structure	6 of 11	55%
Integration of Knowledge & Ideas	3 of 5	60%
Understanding Complex Texts	Proficient	

WRITING 08	
Ideas & Analysis	8
Development & Support	8
Organization	9
Language Use & Conventions	8

If the student took the writing test, the essay was scored on a scale of 1 to 6 by two raters in each of the four writing domains. These domains represent essential skills and abilities that are necessary to meet the writing demands of college and career. The domain scores, ranging from 2 to 12, are a sum of the two raters' scores. The writing score is the average of the student's four domain scores rounded to the nearest whole number. To learn more about the writing score, visit **www.act.org/the-act/writing-scores**.

00000-000000

Information Reported by the Student

College Choice

College Code	9521
Choice	1st

College Selection Items	Rank	
Type	4th	Public 4-year
Student Body Composition	6th	Co-ed
Location	5th	Colorado
Cost (Maximum Tuition)	2nd	$2,000
Size	3rd	5,000–10,000
Field of Study	1st	
Other Factor	7th	

High School Information

Self-Reported Rank	Top Quarter
Self-Reported GPA	3.0–3.4

Major

Accounting
Not Sure / Fairly Sure / Very Sure

Interest–Major Fit
Low / Medium / High

If boxes are not shaded, data was incomplete.

Figure 15.3 The ACT College Report.

ANN C TAYLOR
GRADUATION YEAR: 2020
TEST DATE: APRIL 2019

DOB: AUGUST 22, 2001
GENDER: FEMALE
ACT ID: -54116290
303-337-4850 (C)

7852 W 46TH ST APT 4
WHEAT RIDGE, CO 80033-0234

ANNTAYLOR@EMAIL.COM

HIGH SCHOOL CODE: 061-450
WHEAT RIDGE SR HIGH SCHOOL

The ACT

College Report

Information Reported by the Student

High School Information

Year H.S. graduation or equivalent	2020
Size of senior class	200–399
Type of school	Public
Type of program studied	College Prep

Subjects Studied

Subjects Studied	Years	GPA	AP
English	4	3.7	Yes
Math	4	3.1	No
Social Studies	3	3.5	Yes
Natural Sciences	3	2.9	No
Foreign Language			No
Spanish	1		N/A
German	1		N/A
French	0		N/A
Other	0		N/A

Extracurricular Activities

Extracurricular Activities	High School	College
Instrumental Music	Yes	Yes
Vocal Music	Yes	No
Student Government	Yes	No
Publications	Yes	No
Debate	Yes	Yes
Dramatics	Yes	Yes
Religious Organizations	No	No
Racial/Ethnic Organizations	No	No
Varsity Athletics	Yes	No
Political Organizations	Yes	Yes
Radio/TV	No	No
Fraternity/Sorority/Social	No	No
Service Organizations	No	No

Background Information

Racial/Ethnic background	White
Religious preference	Prefer not to respond
English most frequently spoken in home	Yes
Mother/Guardian 1 ed. level	Bachelor's Degree (4 yrs)
Father/Guardian 2 ed. level	Some college, no degree

Financial Aid Information

Plans to seek financial aid	Yes
Needs help to find work	Yes
Hours per week	11–20

Admissions Enrollment Data

Full-time	Yes
Housing plans	Residence Hall
Citizenship status	Yes
Legal resident of mailing address state	Yes
Physical/Learning disability	--

Interests

First-year honors courses	Yes
Independent study	No
ROTC	Yes

Needs Help With

Educational/Occupational plans	Yes
Writing	No
Reading	No
Study skills	No
Math	Yes

Chances of Success at
UNIVERSITY OF OMEGA

Overall GPA

Student Group	Chance ≥ B	Chance ≥ C
First-Time Students	49%	83%
Business Admin.	45%	81%
Liberal Arts	46%	86%
Education	57%	88%
Engineering	38%	79%

Specific Course Grade

Course	Chance ≥ B	Chance ≥ C
English Composition	89%	95%
College Algebra	20%	62%
History	67%	87%
Chemistry	11%	39%
Psychology	72%	92%

Chances of success are reported for ACT Research Services participants.

For more information about Chances of Success and how your institution can participate, please visit **www.act.org/research** or email **research.services@act.org**.

Figure 15.3 The ACT College Report—continued.

Information Reported by the Student. Student-reported information related to college and career planning on the College Report (see Figure 15.3) includes rankings of college selection criteria, high school rank and range of grade point average, high school information, subjects studied, extracurricular activities, personal background information, financial aid information, admissions enrollment data, and interests in special college programs.

Chances of Success. This section contains two subsections, both of which estimate a student's chances for success at the institution. The first subsection, Overall GPA, refers to the student's estimated chances of earning a first-year college GPA of B or higher and C or higher. The chances are an indication of the level of success a student would be likely to achieve if they were a member of a specified group (e.g., first-time students). The second subsection, Specific Course Grade, contains the student's estimated chances of earning a course grade of B or higher and C or higher in college courses selected by the institution. Results in both subsections are based on a student's ACT score and self-reported high school GPA (described in Chapter 14), and provide information to help determine if a student is academically prepared to succeed at the institution as a member of a certain group or in specific college courses. The predictive information in this section is provided only when the institution participates in the ACT Admissions Service.

References

ACT. (2009). *The ACT interest inventory technical manual*. Retrieved from http://www.act.org/content/dam/act/unsecured/documents/ACTInterestInventoryTechnicalManual.pdf

ACT. (2013). *College choice report part 3: Persistence and transfer*. Iowa City, IA: Author.

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

Chapter 16

ACT State and District Testing

16.1 Overview of State and District Testing

From its inception in 1959 through the 1990s, the ACT test was taken primarily by students interested in pursuing a college degree after high school. Then, in spring 2001, Colorado and Illinois became the first states to offer the ACT to all eleventh graders. Since then, many other states have adopted the ACT statewide as either a census or an optional test. Additionally, individual school districts have opted to provide the ACT to their eleventh-grade students. These states and districts provide an opportunity for their students to take the ACT during the school day and receive college-reportable scores.

In spring 2017, over a million students in 20 states and an additional 900 school districts took the ACT as part of State and District administrations. As shown in Figure 16.1, state-funded administrations of the ACT include census testing in 16 states and four states that provide the ACT as an option to schools. This chapter describes the features and technical characteristics of the State and District administration of the ACT.

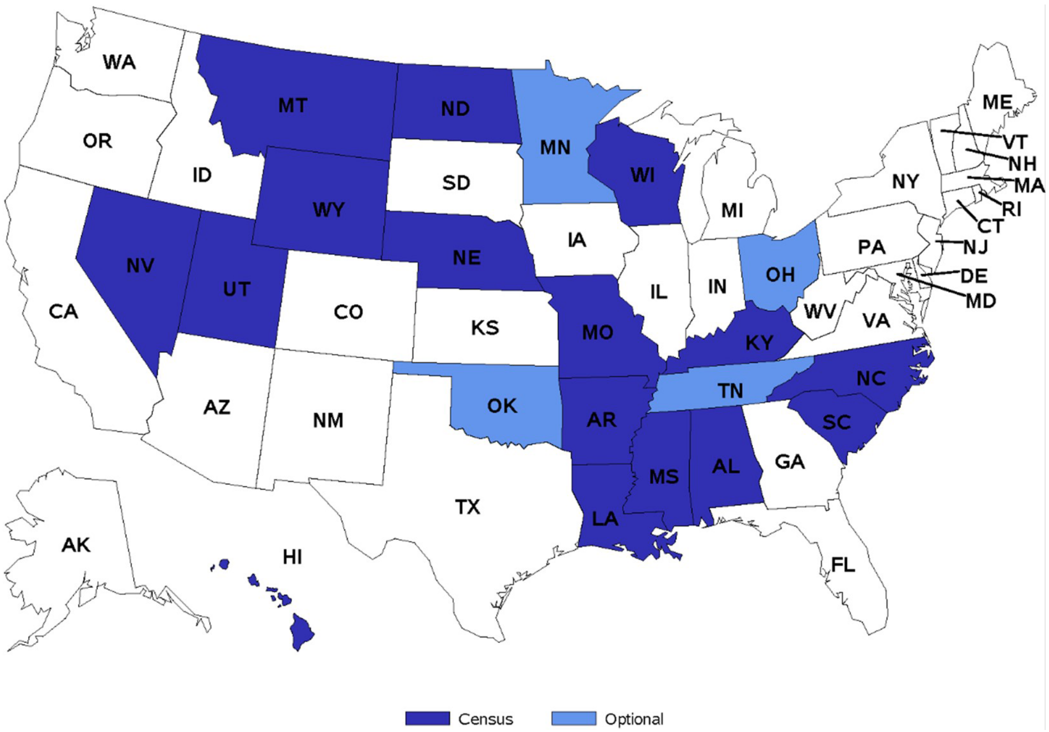


Figure 16.1 States that Administered the ACT in spring 2017.

16.2 Features of State and District Testing

16.2.1 One Test, Multiple Uses

Schools and state departments of education are working with a myriad of competing constraints. These include meeting state and federal accountability requirements around testing students and reporting valid, reliable, and useful scores; working within budget constraints; and balancing the opportunity to learn and classroom instruction time with time spent on test preparation and administration. Given these competing priorities, states are looking for assessments and assessment services that can meet their needs. Using the ACT as a statewide assessment for accountability provides schools with a unique opportunity to fulfill multiple requirements with a single test.

In addition to testing requirements at the state level, the No Child Left Behind (NCLB) Act of 2001 required states to test all students in reading/language arts and mathematics in Grades 3–8 and once in high school (Grades 10–12). It also required testing science once in each of three grade spans: 3–5, 6–9, and 10–12. In 2015, the Every Student Succeeds Act (ESSA) replaced NCLB but continues to require annual testing at the same grade levels and for the same subjects. ESSA also allows for the use of college admissions tests in federal accountability measures. Because the ACT measures English, mathematics, reading, science, and writing, it can be used to fulfill federal testing requirements tied to accountability.

ACT scores can be used to support both standards-based interpretations and norm-referenced interpretations. The ACT College Readiness Benchmarks are scores on the ACT that represent the level of performance required for students to have a 50% chance of obtaining a B or higher or a 75% chance of obtaining a C or higher in corresponding first-year credit-bearing college courses. User norms are also reported both nationally and at the state level so that schools and students can see how their scores compare to the scores of other ACT test takers. More information about ACT norms and ACT Benchmarks can be found in Chapters 7 and 8, respectively.

ACT scores can be used to inform both high school and postsecondary decisions. Because the ACT is curriculum based, ACT scores can be used to inform curriculum decisions and create data-driven intervention strategies. Schools also receive information based on the ACT that can help them better assist their students with postsecondary advising about educational and career planning. Students can use their scores to help inform their postsecondary education plans. With one assessment, needs of both schools and students are met. Among students who participated in a state or district testing program, 62% use the state or district test as their sole ACT test (Allen, 2015a). Moreover, because the students can use their test scores for college and career decisions, they may be more motivated when taking the ACT than they would be when taking other state high school assessments. In addition, students taking the ACT as part of State and District testing can:

- feel less stress due to testing during the school day in a familiar environment
- check off a major part of the college application process
- build confidence in their knowledge and learn about where they still need to improve
- receive personalized information to explore future college and career decisions based on their strengths and interests
- use scores for financial aid and scholarship applications
- provide information to identify college admissions and scholarship opportunities

16.2.2 Opportunity and Inclusion

Prior to the implementation of State and District testing, the ACT was taken by students who intended to go to college after they graduated from high school. These students tended to be higher performing students and students with the means to pay for the test or to obtain a voucher, navigate the registration process, and manage the logistics of showing up on a weekend test date. Administering the ACT during the school day, with no additional cost to students, has provided many students with an opportunity that they would not have otherwise had. Males, African Americans, American Indians, and Hispanic/Latino students, as well as students with lower family income and students whose parents did not attend college are included at higher rates when the ACT is administered by schools (Allen, 2015b). These students, who may not have considered going to college after high school, might reconsider once they receive their ACT scores.

The experience of taking the ACT can help students realize they have the skills to perform college-level course work and give districts the information they need to guide students toward college readiness. The test raises college awareness and exposure among all students.

16.2.3 Flexibility and Convenience

For students, taking the ACT at school during the school day is convenient. Students (and their parents) do not have to worry about test scheduling, as all the logistics are handled by the school. For schools, the logistics involved with administering the ACT are similar to those associated with other standardized tests, and flexibility has been built into the administration procedures to facilitate the process.

ACT provides a choice of test dates for states and districts. There are also test dates provided for makeup, accommodated, and emergency tests. This provides flexibility in determining the test dates that work best for states and districts based on their school calendar.

The ACT is offered in both paper and online formats. For states or districts who choose an online administration format, a testing window is provided so that schools do not have to test every student on the same day if they are limited by technology resources. Materials are provided to schools based on the administration format they choose.

States that use the ACT as part of their federal accountability requirements need to be able to test all their students. Some of these students will need to be tested with accommodations. ACT has a list of allowable accommodations that can be used without invalidating the test score for college admissions purposes. Some students may require additional accommodations that have been approved by the state. These students can still take the ACT, and they can use additional accommodations.

16.2.4 ACT's Services

ACT manages all aspects of the testing process, from the development of test items and forms, to the management of test delivery, scoring, and reporting. In addition, ACT provides a number of services to facilitate score interpretation and use. These services include score interpretation guides, empirical research, technical manuals and reports, and customer support. For states that administer the ACT statewide, additional services are often provided. ACT has partnered with several states to develop materials for submission in the federal peer review process. ACT has also developed state-specific technical reports that include analyses based on data from a particular state. This manual, along with a state-specific technical manual, often provides a majority of the evidence required during the peer review process.

States are also provided with state-specific summaries in the *Condition of College and Career Readiness Report*. For example, Alabama's summary for 2016 included score trend information from 2012 to 2016, as well as comparisons of the state's performance to national performance (http://www.act.org/content/dam/act/unsecured/documents/state01_Alabama_Web_Secured.pdf). This information can help states track participation and performance on the ACT over time and see where their state falls in comparison to the ACT Benchmarks.

16.3 Technical Characteristics of State and District Testing

16.3.1 Test Content

The ACT test forms administered for State and District testing are built to the same content and statistical specifications as the ACT forms administered during National administration dates. The same test development process (described in Chapter 2) is used for both National and State and District testing. The content of the ACT is closely tied to the curriculum of most states and districts because it is developed to reflect what students are learning in school and the postsecondary skills they will need.

States that are considering using the ACT as a measure of English language arts, mathematics, and science must evaluate the alignment of the ACT with state standards. Alignment refers to the content similarity between the education standards a state has adopted and the annual assessments its students take so their progress toward meeting those standards can be measured and evaluated. The ACT is explicitly designed and has been empirically validated to assess student progress toward college and career readiness.

The ACT National Curriculum Survey, conducted every three to four years since 1976, identifies what postsecondary faculty, including instructors of entry-level college and workforce-training courses, expect from their entering students—that is, the knowledge and skills students need to demonstrate to be ready for entry-level postsecondary courses and jobs. ACT then compares these expectations to what is really happening in elementary, middle, and high school classrooms. ACT uses the results of these comparisons to determine the skills and knowledge that should be measured on the ACT and to guide its test blueprints. Therefore, the ACT is an effective way for states that have adopted college and career readiness standards to measure the progress of their students toward meeting those standards.

Many states have found that the ACT is sufficiently aligned with their content standards to administer as an "off-the-shelf" test—that is, without having to change anything about the test. However, if an alignment study reveals gaps between the standards a state wishes to measure and the content covered by the ACT, an "augmented" ACT solution can be considered. Augmenting the ACT involves identifying the state content standards that are not covered by the ACT and developing additional items to measure those standards. The extra items are not part of the standard ACT but are developed by ACT or the state and administered with the ACT in a separately timed session. The augmented ACT solution can provide states with a test that meets their state and federal accountability requirements in addition to providing their students with college-reportable scores.

16.3.2 Administration Procedures

Chapter 5 provides detailed information about the ACT administration procedures. To maintain comparability of experience between National and State and District administrations of the ACT, the administration procedures are very similar. The main difference is that for State and District testing, the ACT is administered during the school day to students at the school, whereas National test administrations occur on the weekend at various testing centers and may not be at a student's school.

The administration dates also differ. Additionally, students who take the ACT at their school may be tested using a computer instead of on paper. Currently all National administrations of the ACT use the paper format. Approximately 10% of students tested as part of State and District testing have an online administration.

Another difference between National administrations and State and District administrations of the ACT is the participants. For State and District administrations, nearly all eleventh-grade students are tested. This includes a number of student groups who do not typically take college admissions tests, including special populations. States are provided with more latitude to assess these populations with the ACT using both approved and nonapproved test accommodations. In the case of nonapproved accommodations, students are able to receive ACT scores. Although these scores are not college-reportable because the accommodations might alter the construct that is being measured, the scores can provide students and teachers with valuable and instructionally actionable information about student performance.

16.3.3 Scales and Scores

For states that administer the ACT as an off-the-shelf solution, they are provided with standard ACT score reports. These score reports contain scores on the ACT multiple-choice tests, STEM, ELA, and Composite scores, as well as a variety of other scores, indicators, and interpretive information to help score users understand student performance and postsecondary interests. More information about the ACT scales, scores, and reports can be found in Chapters 7 and 15.

For states that choose to augment the ACT to test additional state content, states receive the standard ACT score reports, and they also receive score information for the state-specific score scale. To create the state-specific scale, the ACT items are calibrated with the state items using an item response theory (IRT) model. Reported scores can either be transformed from the underlying theta (IRT) scale or based on classical number correct scoring, depending on state preference. In addition, the ACT Benchmarks can be linked to the state-specific score scale. The scores that map to the ACT Benchmarks can be used as empirical performance standards or to inform standard-setting panelists when determining performance level standards on the state assessment.

When a state adopts the ACT as a statewide census test, its average ACT scores are expected to decrease, as are gaps in ACT participation rates across socio-demographic subgroups (Allen, 2015b). This is an expected trend as the group taking the ACT changes from a selective group consisting primarily of high-ability college-bound students to the larger, more diverse state population. States with lower participation rates before adoption tend to have larger initial score decreases. After the initial decrease following the implementation of statewide administration of the ACT, scores on the ACT tend to increase over time (see Figure 16.2). Smaller states also tend to see more fluctuation in scores across time due to changes in the student population from year to year (i.e., cohort effects). After ACT adoption, greater representation among the ACT-tested cohorts is observed for male, African American, American Indian, and Hispanic/Latino students. Students with lower family income and parents who did not attend college are also more likely to be represented after statewide adoption.

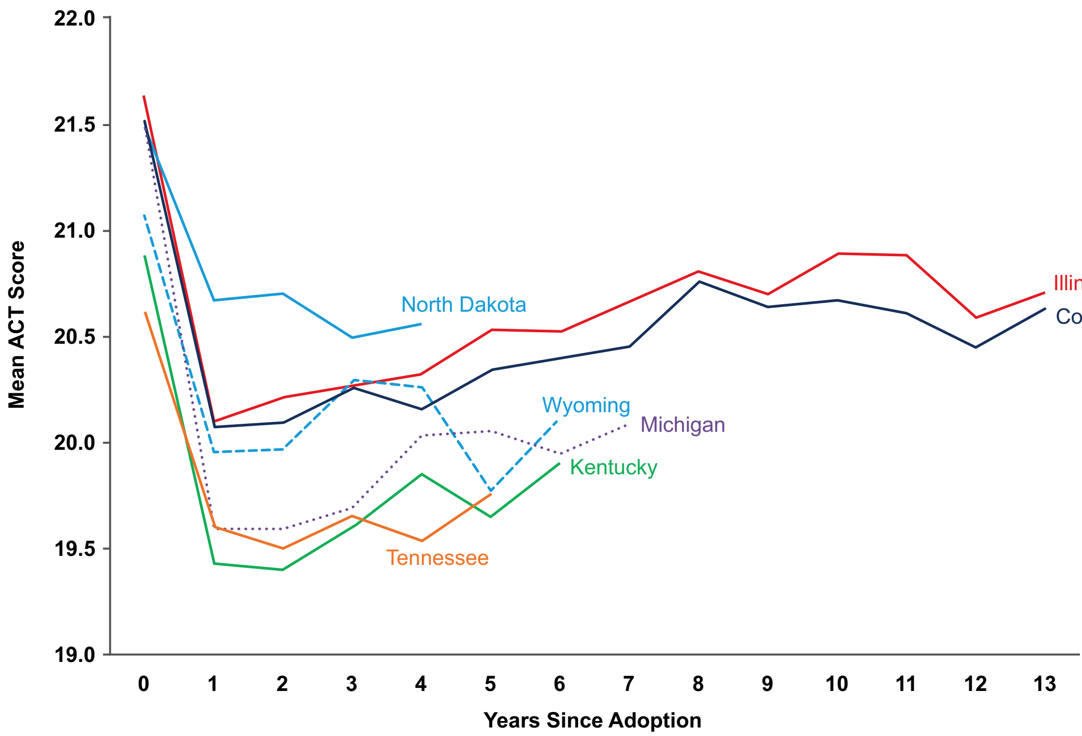


Figure 16.2 Average state ACT Composite score over time.

16.3.4 Standard Setting and Performance Level Descriptors

The ACT multiple-choice tests each have a College Readiness Benchmark that was determined using the predictive relationship between ACT scores and performance in common first-year credit-bearing college courses. Detailed information about how these Benchmarks, or cut scores, were determined is provided in Chapter 8. States that provide the ACT for state accountability purposes or to increase access to postsecondary education for their students may find the single readiness Benchmark sufficient. However, for states that choose to use ACT scores for federal accountability purposes, ESSA requires at least three performance levels (i.e., at least two cut scores).

ACT has supported states in developing additional cut scores, using an empirical standard setting method (Camera, Allen, Moore, 2017). Through this process, ACT collects and summarizes several sources of data including probabilities of success in relevant college courses (Allen, Radunzel, & Moore, 2017), national and state-level impact data, the standard error of measurement for scores on each test, and comparative data points including previous impact data and pass rates on state and NAEP assessments. States who use an empirical standard setting approach for the ACT must consider these three things:

- What are the appropriate outcomes (e.g., common first-year college course grades)?
- What are the criteria of success (e.g., earning a grade of B or higher in a first-year college course)?
- What should the probability of success be (e.g., 50%, 60%, 70%)?

This is a very different task from the usual content-based standard setting approach where judgments are about how likely a “proficient” student would be to get an item correct. ACT’s College Readiness Benchmarks are the scores associated with a 50% chance (the probability of success) of obtaining a B or higher (the criterion of success) in a corresponding first-year credit-bearing college-level course (the outcome) in each subject area.

ACT recommends using the existing ACT College Readiness Benchmarks as a proficient or “passing” cut score. Using the ACT College Readiness Benchmarks will often result in a lower pass rate for states compared to pass rates observed with previous state assessments and performance standards. This does not indicate that performance has decreased, only that the cut scores are more rigorous. Furthermore, if preparing students for college and career readiness is the goal of K–12 education, then aligning state performance standards with college readiness should be a major consideration during standard setting.

To provide a recommendation for basic and advanced cut scores, ACT evaluated all empirical results, including scores on the ACT that linked to the basic and advanced cut scores on the tenth grade Aspire assessment. Based on these data, ACT recommends using a modified confidence interval approach. Using this approach, the basic and advanced cut scores are at least 4 score points apart from the proficient cut score, which is two times the standard error of measurement. These cuts were then evaluated using the probabilities of success in relevant college courses and adjusted down for the basic cut or up for the advanced cut, as needed, to create a fairly consistent probability of success across subjects. The resulting cut scores are similar across subject areas with respect to the probabilities of success in first-year credit-bearing college courses. These recommended performance level cut scores for Basic, Proficient, and Advanced performance, along with the probability of a B or higher in relevant first-year college courses are provided in Table 16.1. Performance level cut scores and probabilities of success are also included for the ACT ELA score, which is based on scores from the ACT English, reading, and writing tests, as ELA is often the ACT score used for accountability purposes.

Table 16.1 Suggested ACT Cut Scores Based on Empirical Standard Setting Results

	Performance level cut scores (probability of success)		
Subject	Basic	Proficient	Advanced
English	13 (0.36)	18 (0.52)	23 (0.68)
Mathematics	18 (0.30)	22 (0.49)	26 (0.69)
Reading	18 (0.36)	22 (0.50)	27 (0.66)
Science	19 (0.30)	23 (0.49)	27 (0.69)
ELA	15 (0.33)	20 (0.52)	24 (0.68)

Some states have already adopted the cut scores recommended in Table 16.1. ACT has also partnered with states to convene panels of stakeholders from across the state to review all the data and determine whether additional adjustments are needed. A content-based and empirical standard setting approach could be combined. In this scenario, standard-setting panelists would make judgments about what students know and can do using both test content and empirical data. Although this approach seems appealing because panelists have both content and data to inform their decisions, it may result in cut scores on the ACT that differ from state-to-state or that are inconsistent with the benchmarks colleges use for admissions decisions.

For states that choose to augment the ACT with additional items to better align the test with their state content standards, empirical, content-based, and hybrid standard-setting approaches are all still possible. For augmented ACT tests, the standard-setting process will need to result in cut scores on the state-specific score scale that includes both ACT and augmented items. One option would be to link the empirical cut scores provided in Table 16.1 to the state scale. This could be done using a single-group design because students will be taking both the ACT and the augmented test. The link between the two scales could be developed using an equipercentile approach, a regression approach, or an IRT approach. The methods would be expected to yield similar results. Another option would be to use the link between the two scales to inform standard- setting panelists. The panelists could make a content-based judgment informed by the relationship between the recommended ACT cut scores and the augmented scale, or they could make an empirically-based judgment where additional sources of data are included in the standard-setting process.

Performance-level descriptors (alternatively known as achievement-level descriptors) describe what a student is expected to know and be able to do at a particular performance level. States who use an empirical standard-setting approach will have empirical performance-level descriptors. As an example, empirical performance-level descriptors are provided below for English:

- Students meeting the Proficient standard in English have a 52% chance of earning a B or higher in a first-year credit-bearing college course in English and a 79% chance of earning a C or higher in the course.
- Students meeting the Basic standard in English have a 36% chance of earning a B or higher in a first-year credit-bearing college course in English and a 70% chance of earning a C or higher in the course.
- Students meeting the Advanced standard in English have a 68% chance of earning a B or higher in a first-year credit-bearing college course in English and an 86% chance of earning a C or higher in the course.

Teachers, parents, and students may want more information about what students at each of the performance levels know and can do. Content-based performance-level descriptors can be created to supply this information. Whereas for content-based standard settings, the content-based performance-level descriptors are developed *before* the standard setting and provide the basis for standard-setting judgments, content-based performance-level descriptors are created *after* an empirical standard setting. One method for creating content-based performance-level descriptors after an empirical standard setting is to use a scale-anchoring approach (Sinharay, Haberman, & Lee, 2011). Scale anchoring involves reviewing score distributions for ACT items and identifying items that have been answered correctly by 80% or more of test takers within a performance level. These items measure the knowledge and skills that test takers of a particular performance level appear to have mastered. Once a set of items is identified for each performance level, content experts summarize the knowledge and skills that test takers at a particular performance level are able to demonstrate. A panel reviews and revises these performance-level descriptors to clarify wording, verify continuity across the levels, and ensure adequate representation of the state content standards.

The scale-anchoring process has been used by ACT during the development of the ACT College and Career Readiness Standards. Through the scale-anchoring process, ACT has identified the set of knowledge and skills examinees possess at various score ranges on the ACT. For the four multiple-choice tests, content-based performance-level descriptors are provided for six scale score ranges: 13–15, 16–19, 20–23, 24–27, 28–32, and 33–36. For writing, content-based performance-level descriptors are provided for five score ranges: 3–4, 5–6, 7–8, 9–10, and 11–12. The ACT College and Career Readiness Standards can provide a well-established starting point for the development of state content-based performance standards. More information about the ACT College and Career Readiness Standards is provided in Chapter 8.

States that choose a content-based or hybrid standard-setting method will have to develop content-based performance-level descriptors prior to the standard setting meeting. Although a scale anchoring approach to performance-level descriptor development would not be needed, states can still benefit from the content-based performance-level descriptors already contained in the ACT College and Career Readiness Standards. After a content-based standard setting, ACT can provide states with empirical performance-level descriptors using the resulting cut scores.

16.3.5 Equating and Mode Comparability

Equating procedures are used to maintain comparability of scores across test forms. Both the National and State and District forms of the ACT use the same equating procedures, as described in detail in Chapter 9. Augmented ACT test forms are equated to previous forms using IRT equating procedures, as vetted and approved by the state. ACT also verifies the equating through classical test theory methods, as appropriate. In addition, the ACT scale scores for the off-the-shelf program can be used as external anchors to link the augmented test scores, which provides another check on the stability of the augmented test scores over time. ACT has found using multiple methodologies an effective way to ensure the stability of reported scores over time.

Because the ACT is offered in both paper and online testing modes, mode comparability studies have been conducted to verify that the level of difficulty of the ACT is the same across modes. Chapter 12 provides a detailed description of the studies that have been conducted to address mode comparability for the ACT.

16.3.6 Reliability and Validity

The reliability and validity of the ACT scores are described in detail in Chapters 10 and 11. ACT scores are comparable across National and State and District administrations and the reliability and validity information described in Chapters 10 and 11 applies to all ACT scores. For states that have adopted the ACT statewide, a state-specific technical report can be developed that includes additional reliability and validity information using data from the state's student population. Additionally, for states that choose an augmented ACT solution, it is important to provide the reliability and validity evidence for scores on the state score scale. With additional items, the reliability of the subject test scores can be expected to increase. However, since the augmented items are designed to measure content that differs from what is on the ACT, the predictive validity of the state scores may not be as high as the predictive validity of the ACT for measuring college and career readiness. Validity evidence for the state should be collected based on the score interpretations the state hopes to make from the augmented ACT scores.

References

Allen, J. (2015a). How many students use ACT State and District testing to take their sole ACT test? *ACT Data Byte 2015-9*. Iowa City, IA: ACT, Inc.

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

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, Inc.

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.

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

Appendix A:

The 1995 ACT National Norming Study

This appendix discusses the special study conducted in 1995 for norming the ACT tests.

Norming and Score Scale Data

In October 1988, ACT conducted a national study involving more than 100,000 high school students (ACT, 1997). This study, called the Academic Skills Study, provided the data that ACT used to revise the ACT score scales and to provide nationally representative norms. In October 1995, ACT conducted another national study, this one involving over 24,000 high school students. Data from the 1995 study were used to examine the score scale and to update the national norms given the use of calculators on the mathematics test. The 1995 study is discussed in this appendix mainly in regard to updating the norms for the ACT tests.

Sampling for the 1995 Study

In the sample used for obtaining new norms for the ACT, one form of the calculator-not-permitted ACT and two forms of the calculator-permitted ACT were administered to twelfth graders. All three forms were administered at each school to randomly equivalent groups of examinees, using spiraling. The booklets were spiraled within classroom, meaning that some students were allowed to use a calculator on the mathematics test while other students were not.

Sample Design and Data Collection. The target population consisted of students enrolled in twelfth grade in schools in the US The target population included students in both private and public schools. The sample size was chosen with the goal of achieving a precision level that would enable estimating any probability to within .05 with probability .95. The sample was explicitly stratified by region and school size. It was further implicitly stratified by affiliation and the percentage of minority students. A systematic sample was selected from each stratum. (Harris, 1997, offers more information on the sampling).

In anticipation that some schools would not participate in the study, many more schools were invited to participate than were required to achieve the targeted precision. During the recruitment the number of participating schools in each stratum was carefully monitored, so as to maintain the representativeness of the sample with respect to the stratification variables. In addition, a backup sample was chosen so that additional schools could be chosen from strata for which there were too few schools agreeing to participate. Schools were asked to test all students in each grade. A few schools were allowed to administer the spiraled test batteries to randomly selected subsamples of their students. Makeup testing for students who were absent was strongly encouraged.

Response Rates. One type of nonresponse in this study was among schools: not every school invited to participate did so. Attempts were made to choose the replacement schools from the same strata as the schools they were replacing so that the obtained sample would be representative with respect to the stratification variables. Nevertheless, it is conceivable that a school's willingness to participate in this study could be related to their students' academic development, independently of these variables. If this were true, then the nonresponse among schools would introduce a bias in the results. It is not believed the selection of schools had any significant biasing effect in computing the norms.

A second type of nonresponse was among students within a participating school. One source of student nonresponse was absenteeism (schools were encouraged to retest students who were absent). Within-school student participation rates were monitored, and those schools with response rates less than 50% were contacted by phone. If there was not a reasonable justification for the less-than-50% response rate—such as that the school chose to test only a randomly selected subsample of students—the school was eliminated from further analyses. Four schools were deleted for this reason. It is believed that for the sample as a whole, student nonresponse did not have any important biasing effect.

Data Editing. Data from two schools were eliminated due to irregularities in the administration of the tests. From the 67 remaining schools, examinees with problematic records were excluded (e.g., grade level not determinable, test form not determinable, zero raw score on one of the four tests, over two-thirds of the items omitted on any of the four tests). A minimal number of returned answer sheets were excluded. Final sample sizes for all examinees (national) and the subset of examinees who indicated they were college-bound are shown in Table A.1. A college-bound student was defined as a student who indicated he or she was planning to attend a two- or four-year college or university after high school graduation.

Table A.1 Examinee Sample Sizes for Updating the Norms (1995 Study)

Grade	National	College-Bound
12	2,981	2,356

Weighting. For the norming process, individual examinee records were multiplied by weights. Weighting is an adjustment performed to match the characteristics of the sample to that of the target population. This is done by either increasing or decreasing the importance of a particular observation, depending on the stratum where the observation is located. The result of this process is that the weighted sample will have proportions in each stratum equal to the proportions in each stratum in the population. For purposes of weighting and calculating standard errors, any stratum with fewer than two schools was combined with another stratum. In addition, weights were truncated. This was done so that no one school or student score would have an undue influence on the results. Harris (1997) provides details on the procedure used to determine the weights.

Sample Representativeness. The representativeness of the sample is a consequence of the relative levels of success in recruiting schools of different sizes and from different parts of the country and having these schools test their entire twelfth-grade class. One way to determine the character and extent of sample bias is to compare the demographic characteristics of the sample of examinees with the US statistics for various demographic variables presented in Table A.2. Precisely comparable US data for the population of interest were not available. However, the available data allowed for a general examination of the representativeness of the sample with respect to the demographic variables.

As indicated in Table A.2, the weighted sample appears to be reasonably representative of the national population. The actual discrepancy between African American students and male students is probably considerably less than appears in the table for two reasons. First, some students did not respond to the question concerning racial/ethnic background, or chose “other” or “prefer not to respond” as their response. Second, the US percentages in Table A.2 are based on students in Grades K–12, not just Grade 12. To the extent that African American students and male students drop out at higher rates than other students, the US percentage will be overstated. Even though region was used as a stratification variable, these percentages are also slightly different from the national percentages. This is due to the truncation of the weights previously mentioned.

Obtained Precision. The targeted precision level was to estimate any probability to within .05 with probability .95. The actual obtained level of precision for the norms was estimation of any probability to within .12 with probability .95. This is far from the targeted value for two reasons. First, fewer schools were available for analysis than had been targeted. Second, among those schools that did participate, there was an unusual amount of homogeneity within a school. That is, the students within a school were far more similar than was expected. Students within each school all tended to do well or all tended to do poorly. This phenomenon reduced the efficiency of the sample.

Table A.2 Selected Demographic and Educational Characteristics for the 1995 Norming Study Sample

Category identifier used in study	Weighted sample proportion Grade 12	US proportion*	US category identifier
Gender			
Female	.55	.49	Female
Male	.45	.51	Male
Racial/Ethnic Origin			
African American/Black	.12	.17	Black
American Indian, Alaska Native	.01	.01	Indian
Caucasian American/White	.70	.66	White
Mexican American/Chicano	.03	.13	Spanish Origin
Puerto Rican, Cuban, Other Hispanic	.03	—	—
Asian American/Pacific Islander	.02	.04	Asian
Multiracial	.01	—	—
Other, Prefer Not to Respond, Blank	.06	—	—
School Affiliation			
Private	.05	.08	Private
Public	.95	.92	Public
Geographic Region			
East	.38	.44	East
Midwest	.28	.19	Midwest
Southwest	.16	.13	Southwest
West	.18	.24	West

*US proportions obtained from United States Department of Education, *Digest of Education Statistics 1996* (pp. 23 and 60).

Norms for the National Sample

The norms for the ACT are intended to represent the national population of twelfth-grade students and the national subpopulation of twelfth-grade students who reported that they plan to attend a two- or four-year college when tested at the beginning of twelfth grade. The norms were obtained from the 1995 nationally representative sample using the weighting procedures described previously. After data editing, examinees who were allowed to use a calculator on the mathematics test were included to produce the norms.

Data from the national sample were used to develop cumulative percentages (percents-at-or-below) for each ACT test score, the Composite score, and the subscores. The percent-at-or-below corresponding to a scale score is defined as the percentage of twelfth-grade examinees with scores equal to or less than that scale score.

Tables A.3 and A.4 contain percents-at-or-below for the four ACT multiple-choice test scores and the Composite score for twelfth-grade college-bound and national examinees, respectively. Calculators were allowed on the mathematics test beginning on the October 1996 test date. The norms reported in Tables A.3 and A.4 for the Composite scores and mathematics test scores are not appropriate for the ACT taken prior to October 1996.

Table A.3 ACT Norms for College-Bound High School Students (Cumulative Percentages for Test Scale Scores Based on the 1995 Norming Study)

	Percent at or below				
Scale Score	English	Mathematics	Reading	Science	Composite
1	01	01	01	01	01
2	01	01	01	01	01
3	01	01	01	01	01
4	01	01	01	01	01
5	01	01	01	01	01
6	01	01	01	01	01
7	01	01	01	01	01
8	03	01	02	01	01
9	05	01	04	01	01
10	10	01	08	03	01
11	16	01	14	07	04
12	20	06	21	11	09
13	26	09	29	23	17
14	32	19	34	28	26
15	37	30	39	38	34
16	43	40	45	43	42
17	49	49	50	51	50
18	56	58	57	61	56

Table A.3 ACT Norms for College-Bound High School Students (Cumulative Percentages for Test Scale Scores Based on the 1995 Norming Study)—continued

	Percent at or below				
Scale Score	English	Mathematics	Reading	Science	Composite
19	62	64	60	69	62
20	66	70	66	74	68
21	71	76	71	78	74
22	75	81	73	83	80
23	79	83	77	88	84
24	83	87	81	91	88
25	86	90	83	95	91
26	90	93	86	96	94
27	93	95	90	97	96
28	96	97	92	98	97
29	97	98	94	99	98
30	98	99	95	99	99
31	99	99	97	99	99
32	99	99	97	99	99
33	99	99	98	99	99
34	99	99	99	99	99
35	99	99	99	99	99
36	99	99	99	99	99

Table A.4 ACT Norms for National High School Students (Cumulative Percentages for Test Scale Scores Based on the 1995 Norming Study)

	Percent at or below				
Scale Score	English	Mathematics	Reading	Science	Composite
1	01	01	01	01	01
2	01	01	01	01	01
3	01	01	01	01	01
4	01	01	01	01	01
5	01	01	01	01	01
6	01	01	01	01	01
7	02	01	01	01	01
8	04	01	03	01	01
9	08	01	05	02	01
10	14	01	09	04	02
11	20	02	17	10	05
12	25	07	25	15	13
13	31	13	34	26	23
14	39	24	39	33	33
15	44	35	45	43	41
16	50	46	51	49	49
17	56	56	56	56	57
18	62	64	62	66	62
19	68	69	65	73	68
20	71	75	70	78	73
21	76	79	75	82	78
22	79	84	77	86	83
23	83	86	80	90	87
24	86	89	84	92	90
25	89	92	86	95	93
26	92	94	88	97	95
27	94	96	91	97	97
28	96	97	93	98	98

Table A.4 ACT Norms for National High School Students (Cumulative Percentages for Test Scale Scores Based on the 1995 Norming Study)—continued

	Percent at or below				
Scale Score	English	Mathematics	Reading	Science	Composite
29	98	98	95	99	98
30	99	99	96	99	99
31	99	99	97	99	99
32	99	99	98	99	99
33	99	99	98	99	99
34	99	99	99	99	99
35	99	99	99	99	99
36	99	99	99	99	99

References

ACT. (1997). *ACT Assessment technical manual*. Iowa City, IA: Author.

Harris, D. J. (Ed.). (1997). *Methodology and results from the Fall, 1995 Calculator Study for ACT and PLAN (including EXPLORE norming)*. Iowa City, IA: ACT, Inc.