Wisconsin Technical Manual
2019–2020

Prepared by: ACT
Date: June 24, 2021

Preface

Wisconsin 2019–2020

Preface

The purpose of this manual is to supplement the additional information that can be found in *the ACT® Technical Manual*, which provides technical information about the ACT test, including national-level reliability, scaling and equating, and validity evidences. This technical report provides Wisconsin-specific information based on the 2019–2020 academic year.

The ACT Technical Manual describes various content and psychometric aspects of the ACT and documents the collection of validity evidence that supports appropriate interpretations of test scores. 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.

Table of Contents

Chapter 1

Introduction

Chapter 1: Introduction

1.1 ACT State 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. Beginning in the spring of 2001, states began to offer the ACT to 11th graders in preparation for college admissions processes. Since then, states have adopted the ACT as either a statewide census or an optional test. Additionally, individual school districts have opted to provide the ACT to their 11th-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.

During the 2019–2020 academic year, the state of Wisconsin provided 11th-grade students with the opportunity to take the ACT test during regular school hours at schools certified as ACT state testing sites. Students taking the ACT under standard time and with ACT-approved accommodations are able to use their scores when applying for college admissions. Information pertaining to administration of the ACT in a specific state can be found in the portal for each state. Wisconsin-specific information and notifications can be found at the website below:

http://www.act.org/content/act/en/products-and-services/state-and-district-solutions/wisconsin.html

1.2 Proposed Uses of Test Scores

ACT test data 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. States, such as Wisconsin, use the ACT as part of their statewide assessment and accountability. Postsecondary institutions use ACT results for admission and course placement decisions. 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.

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 it 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.

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. In addition, students taking the ACT as part of Wisconsin's state participation may:

- 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

1.3 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. The experience of taking the ACT can help students realize they have the skills to perform college-level coursework and give districts the information they need to guide students toward college readiness.

For students, taking the ACT at school during the school day is convenient. Students 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 including 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.

Each academic year, WDPI organizes the administration dates with ACT to identify a three 2-week testing windows during which Wisconsin students are able to take the ACT during the school day. The initial testing window is designated for all students to be able to test and is referred to as the "primary" test form in subsequent chapters of the technical report. Additional testing days are designated as makeup and emergency dates for students who were unable to participate in the initial statewide testing administration.

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 approved accommodations that can be used without invalidating the test score for college admissions purposes.

1.4 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 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 few 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.

As part of the effort to support WI DPI in preparing all students to be college and career ready, ACT and WI DPI developed a research study to evaluate the alignment of the ACT to Wisconsin Academic Standards (WAS). In the fall of 2017, ACS Ventures (ACS) led a review of the Wisconsin standards that showed significant, but not full, alignment of the ACT to Wisconsin's standards. Since the conclusion of this study, ACT's content staff members have been working to improve the overall alignment of the

ACT by ensuring that test items address the content and depth of knowledge targets, as well as providing content coverage across ability levels. Along with other resources for Wisconsin students regarding the ACT, the report produced by ACS can be found on the WI DPI website https://dpi.wi.gov/assessment/act/resources (see the Technical Manuals section of the website). Additional information regarding ACT's College and Career Readiness standards are located at http://www.act.org/content/act/en/college-and-career-readiness/standards.html.

1.5 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 of *the ACT Technical Manual*. Because the ACT can be 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 of *the ACT Technical Manual* provides a detailed description of the studies that have been conducted to address mode comparability for the ACT.

Chapter 2

Test Development

Chapter 2: Test Development

2.1 Overview

This chapter describes ACT's tests and the development process and writing test prompt construction process. A brief overview of 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.

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.

2.2.1 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 utilized 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 2.1 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.

2.2.2 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 3.

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. When constructing test forms, ACT content staff members review and evaluate items and test blueprints at various levels of detail to ensure that the content featured on each form meets highly specific content specifications. Presented in this technical report are high level details that guide the development of the ACT test forms.

Additional details pertaining to the alignment of Wisconsin Academic Standards (WAS) to ACT content standards can be found in the independent research study performed by ACS. The report provides a detailed outline of the processes and tasks performed by subject matter experts to evaluate how well the ACT meets the needs of the WDPI in assessing the college and career readiness of Wisconsin students. The report from ACS can be accessed through the WDPI ACT resources website at https://dpi.wi.gov/assessment/act/resources (see the Technical Manuals section of the website).

Table 2.2 DOK Level Descriptions for Mathematics

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

2.2.3 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 2.3 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.

2.2.4 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 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 2.10). 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. 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 2.4 DOK Level Descriptions for Science

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.

2.2.5 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 tells 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 2.5 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.

2.3 Test Development Procedure

2.3.1 Review of Test Specifications

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

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 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).

2.3.2 Selection of Item and Prompt 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.

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.

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.

2.3.3 Item and Prompt 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.

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.

2.3.4 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.

2.3.5 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.

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. 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.

Prompts are evaluated for 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.

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.

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.

2.3.6 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. The distributions of item difficulty levels obtained on recent forms of the four tests are displayed in Table 2.6. The data in Table 2.6 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 2.6 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 2.6 Difficulty[a] Distributions and Mean Discrimination[b] Indices for ACT Test Items, 2015–2016

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 or mathematics, 40 for reading, and 40 for science.
[d]Completion rate is the proportion of examinees who answered each of the last five items.

2.3.7 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 Test Specifications

2.4.1 English Specifications

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.

Scores and Reporting Categories

1. 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.

2. Knowledge of Language

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

3. 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.

English Test Blueprint

Table 2.7 Specification Ranges by Reporting Category for English

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

2.4.2 Mathematics Specifications

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. 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.

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. 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 2.8 shows the number of items that contribute to each reporting category score. Descriptions of each reporting category follow.

1. 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 non-right 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.

2. 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.

3. 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.

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.

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.

Mathematics Test Blueprint

Table 2.8 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 2.8 Specification Ranges by Reporting Category for Mathematics

Reporting category	Number of items	Percentage of test
Preparing for Higher Mathematics	34-36	57-60%
Number & Quantity	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 Number of Items	60	100%

2.4.3 Reading Specifications

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. A description and the approximate percentage of the test devoted to each reporting category are given below.

1. **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.

2. **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.

3. **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.

Reading Test Blueprints

Table 2.9 Specification Ranges by Reporting Category for Reading

Reporting Category	Number of Items	Percentage of Test
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%

2.4.4 Science Specifications

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 2.10. The overall score on the science test is also used, with the mathematics score, to determine the STEM score.

1. **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).

2. **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).

3. **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).

Science Test Blueprint

Table 2.10 Specification Ranges by Reporting Category for Science

Reporting Category	Number of Items	Percentage of Test
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 2.11 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 Number of Items	6	40	100%

2.4.5 Writing Specifications

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 2–121 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.

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 2.13. If the ratings disagree by more than one point, a third rater evaluates the essay and resolves the discrepancy.

1. **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.

2. **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.

3. **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.

4. **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.

Performance Scoring

Various performance scoring processes and procedures are utilized for scoring the ACT test with writing, 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.

Rater Training and Qualification

The range-finding 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 range-finding 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.

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 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 2.12 provides a sample of some of the available reports.

Table 2.12 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 distributions.

Table 2.13 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, 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	The writer generates an argument that productively engages with multiple perspectives on	Development of ideas and support for claims deepen understanding. A mostly integrated line of purposeful reasoning and	The response exhibits a productive organizational strategy. The response is mostly unified by	The use of language works in service of the argument. Word choice is precise. Sentence structures are

	Ideas & analysis	Development & support	Organization	Language use & conventions
essay.	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, tensions, and/or underlying values and assumptions.	illustration capably conveys the significance of the argument. Qualifications and complications enrich ideas and analysis.	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.	clear and varied often. Stylistic and register choices, including voice and tone, are purposeful and productive. While minor errors in grammar, usage, and mechanics may be present, they do not impede understanding.
Score 4: 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, 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 on 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	The writer generates an argument that responds to	Development of ideas and support for claims are mostly relevant	The response exhibits a basic organizational structure. The	The use of language is basic and only somewhat clear.

	Ideas & analysis	Development & support	Organization	Language use & conventions
some developing skill in writing an argumentative essay	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.	but are overly general or simplistic. Reasoning and illustration largely clarify the argument but may be somewhat repetitious or imprecise.	response largely coheres, with most ideas logically grouped. Transitions between and within paragraphs sometimes clarify the relationships among ideas.	Word choice is general and occasionally imprecise. Sentence structures are usually clear but show little variety. Stylistic and register choices, including voice and tone, are not always appropriate for the rhetorical purpose. Distracting errors in grammar, usage, and mechanics may be present, but they generally do not impede understanding.
Score 2: Responses at this score point 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

	Ideas & analysis	Development & support	Organization	Language use & conventions
				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.

2.5 Internal Structure

This section presents evidence showing the extent to which the scoring and reporting structure of the ACT assessment is consistent with the internal structure reflected in observed data. In each ACT subject test, besides an overall score, subscores associated with content-related reporting categories (reporting category scores) are reported. Table 2.14 provides a blueprint of subscore reporting structure and the target number of items for each reporting category.

Table 2.14 Reporting Categories of the ACT Assessment

Subject	Reporting category	Abbreviation	Target number of items
English	Production of Writing	PoW	22–24
	Knowledge of Language	KLA	11–13
	Conventions of Standard English	CoE	39–41
Math	Preparing for higher math	PHM	34–36
	Number & Quantity	NAQ	4–6
	Algebra	Algebra	7–9
	Functions	Functions	7–9
	Geometry	Geometry	7–9
	Statistics & Probability	SAP	5–7
	Integrating Essential Skills	IES	24–26
	Modeling	Modeling	>=16
Reading	Key Ideas & Details	KID	22–24
	Craft & Structure	CAS	10–12
	Integration of Knowledge & Ideas	IOK	6–7
Science	Interpretation of Data	IOD	18–22
	Scientific Investigation	SIN	8–12
	Evaluation of Models, Inferences & Experimental Results	EMI	10–14

In order to test the hypothesis that the subscore reporting structure is consistent with the internal structure shown in observed data, a series of confirmatory factory analyses (CFA) was conducted. The CFA is an approach to test whether a theoretically driven model of internal structure is consistent with observed data for a test or a measure. For each ACT subject test, an internal structure model was built based on the test blueprint and reporting category classification. For example, the structure model for the science test of the primary form for Wisconsin is depicted in Figure 2.1. In this model, three latent factors are defined, which represent three content-related reporting categories in the science test (i.e. Interpretation of Data, Scientific Investigation, Evaluation of Models, Inferences, and Experimental Results). Each of these latent factors is measured by a certain number of observed variables, which are test items. The three latent factors can be correlated with each other. Each item only measures one latent factor. Similar models were built for English, mathematics, and reading based on the test blueprints. Then, how well the theoretically driven models fit observed data was evaluated through CFAs with the empirical data from Wisconsin spring 2020 ACT tests.

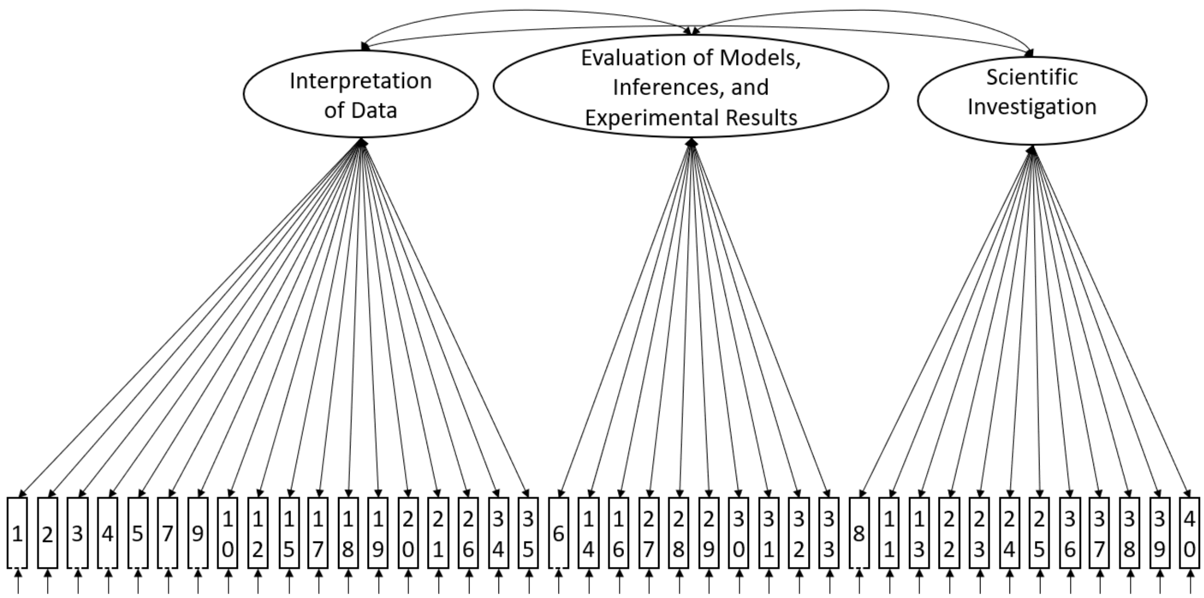

Figure 2.1. Latent factor model for the ACT science test of the primary form for Wisconsin.

The CFA results were evaluated based on model fit statistics and factor loadings of items on latent factors. Table 2.15 presents the model fit statistics for the four subjects examined: English, mathematics, reading, and science. First, a chi square (χ^2) statistic, the most frequently cited index of absolute fit, was examined. The chi square compares the observed covariance matrix with our theoretically proposed covariance matrix. However, the result of a chi square test is commonly known to be greatly influenced by the sample size, so that good-fit models can sometime produce statistically significant chi square test results when sample sizes are large. Thus, additional fit statistics were examined and were mainly relied on in interpreting model fit results. The additional fit statistics include Comparative Fit Index (CFI; Bentler, 1990) and Tucker Lewis Index (TLI; Tucker and Lewis, 1973), which are incremental fit indices and indicate how much better a model fits the data compared to a baseline model where all variables are uncorrelated. Values of the two statistics can range from 0–1. Values above .90 indicate reasonable fit, and values above .95 indicate good model fit (Bentler, 1990; Hu & Bentler, 1999). Another additional fit statistic examined is Root Mean Square Error of Approximation (RMSEA), which is a measure of “discrepancy per degree of freedom” in a model when taking parsimony into account in model comparison. Values less than 0.08 suggest reasonable model fit, and values less than 0.05 suggest good model fit (Browne & Cudeck, 1993). The fit statistics of the internal structure models of English, mathematics, reading, and science are shown in Table 2.15. For all four subjects, all statistics showed reasonable fit to the observed data.

Table 2.15. Fit Statistics of Internal Factor Structure Models of the ACT Tests

	Chi square	DF	*P*-value of Chi-square	RMSEA	CFI	TLI
English	137831.531	2697	0	0.031	0.924	0.922
Mathematics	55152.961	1695	0	0.025	0.954	0.952
Reading	44398.560	737	0	0.034	0.935	0.931
Science	25503.354	737	0	0.025	0.961	0.959

Next, the adequacy of factor structure model was evaluated by investigating the factor loadings. Factor loadings represent the strength of the association between a latent factor and its observed indicators. In the current analysis, factor loadings indicate the strength of association between a reporting category and a certain number of items measuring it. Table 2.16 shows the average of standardized factor loadings of items for each reporting category. For all four subject tests, most of the factor loadings suggest moderate (above 0.4) to strong association (above 0.5) between the items and each of the reporting categories. For details of factor loading of each individual item, see Appendix 1.

Table 2.16 Average Factor Loadings of Items on Latent Factors for the ACT Assessment

<table>
<tr><th colspan="7">English</th></tr>
<tr><td>Reporting categories</td><td colspan="2">Production of Writing</td><td colspan="2">Knowledge of Language</td><td colspan="2">Conventions of Standard English</td></tr>
<tr><td>Average factor loadings</td><td colspan="2">0.504</td><td colspan="2">0.582</td><td colspan="2">0.507</td></tr>
<tr><th colspan="7">Mathematics</th></tr>
<tr><td>Reporting categories</td><td>Number & Quantity</td><td>Algebra</td><td>Functions</td><td>Geometry</td><td>Statistics & Probability</td><td>Integrating Essential Skills</td></tr>
<tr><td>Average factor loadings</td><td>0.511</td><td>0.480</td><td>0.526</td><td>0.497</td><td>0.485</td><td>0.492</td></tr>
<tr><th colspan="7">Reading</th></tr>
<tr><td>Reporting categories</td><td colspan="2">Key Ideas & Details</td><td colspan="2">Craft & Structure</td><td colspan="2">Integration of Knowledge & Ideas</td></tr>
<tr><td>Average factor loadings</td><td colspan="2">0.482</td><td colspan="2">0.525</td><td colspan="2">0.563</td></tr>
<tr><th colspan="7">Science</th></tr>
<tr><td>Reporting categories</td><td colspan="2">Interpretation of Data</td><td colspan="2">Scientific Investigation</td><td colspan="2">Evaluation of Models, Inferences & Experimental Results</td></tr>
<tr><td>Average factor loadings</td><td colspan="2">0.526</td><td colspan="2">0.467</td><td colspan="2">0.521</td></tr>
</table>

Considering all above evidences, the ACT test subscore reporting structure fits the internal structure of test reflected in observed data reasonably well. This further implies that internal structure of the ACT tests is reasonably representing Wisconsin's content standards because of the similarity of subscore structure of ACT tests and Wisconsin's content standards.

2.6 DIF Analysis for Multiple-Choice Items

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.

After each operational administration, differential item functioning (DIF) analysis procedures are conducted on the test data. DIF analyses were conducted for spring 2020 Wisconsin students taking the primary form. DIF can be described as a statistical difference between the probability of the specific population group (the "focal" group) getting the item right and the comparison population group (the "reference/base" group) getting the item right given that both groups have the same level of achievement with respect to the content being tested. The procedures used for the analysis include the standardized difference in proportion-correct (STD) procedure and the Mantel-Haenszel common odds-ratio (MH) procedure. For a description of these statistics and their performance overall in detecting DIF, see the ACT Research Report *Performance of Three Conditional DIF Statistics in Detecting Differential Item Functioning on Simulated Tests* (Spray, 1989).

Both the STD and MH techniques are designed for use with multiple-choice items, and both require data from significant numbers of students to provide reliable results. We adopted minimum sample size requirement based on an industry practice: a minimum of 300 students for focal group and 700 students for total (Zwick, 2012). As a result, DIF analyses were conducted on each multiple-choice item for the initial form on nine group comparisons. The groups compared were Male/Female, White/Asian, White/African American, White/Hispanic, White/American Indian and White/Two or more races.

Using MH procedure, items with MH odd ratio values smaller than 0.5 or larger than 2.0 were flagged. Using STD procedure, items were flagged when the values of STD were higher than 0.10. Of the 215 items on the ACT assessment, only five items were flagged for C-level DIF based on the data of the students in Wisconsin.

Table 2.17 DIF Results Based on Wisconsin Student Data

Subject	DIF Code	Comparison
English	C+	Female vs Male
English	C-	Asian vs White
English	C+	Black vs White
English	C+	Asian vs White
English	C-	Female vs Male

References

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological bulletin, 107*(2), 238.

Browne, M. W., & Cudeck, R. (1992). Alternative ways of assessing model fit. *Sociological Methods & Research, 23*(2), 230–258.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural equation modeling: a multidisciplinary journal, 6*(1), 1–55.

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

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

Tucker, L. R., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika, 38*(1), 1–10.

Webb, N. L. (2002, March 28th). *Depth-of-knowledge levels for four content areas*. Unpublished manuscript. Retrieved from: http://ossucurr.pbworks.com/w/file/fetch/49691156/Norm%20web%20dok%20by%20subject%20area.pdf

Zwick, R. (2012). *A review of ETS differential item functioning assessment procedures: Flagging rules, minimum sample size requirements, and criterion refinement* (Research Report No. 1208). Princeton, NJ: Educational Testing Service.

Chapter 3

Accessibility

Chapter 3: Accommodations

3.1 Accommodations

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 they should be doing independently; the supports just make interaction and communication possible and fair for each examinee.

3.1.1 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 3.1 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 3.1, depending on whether the support must be ordered in advance).

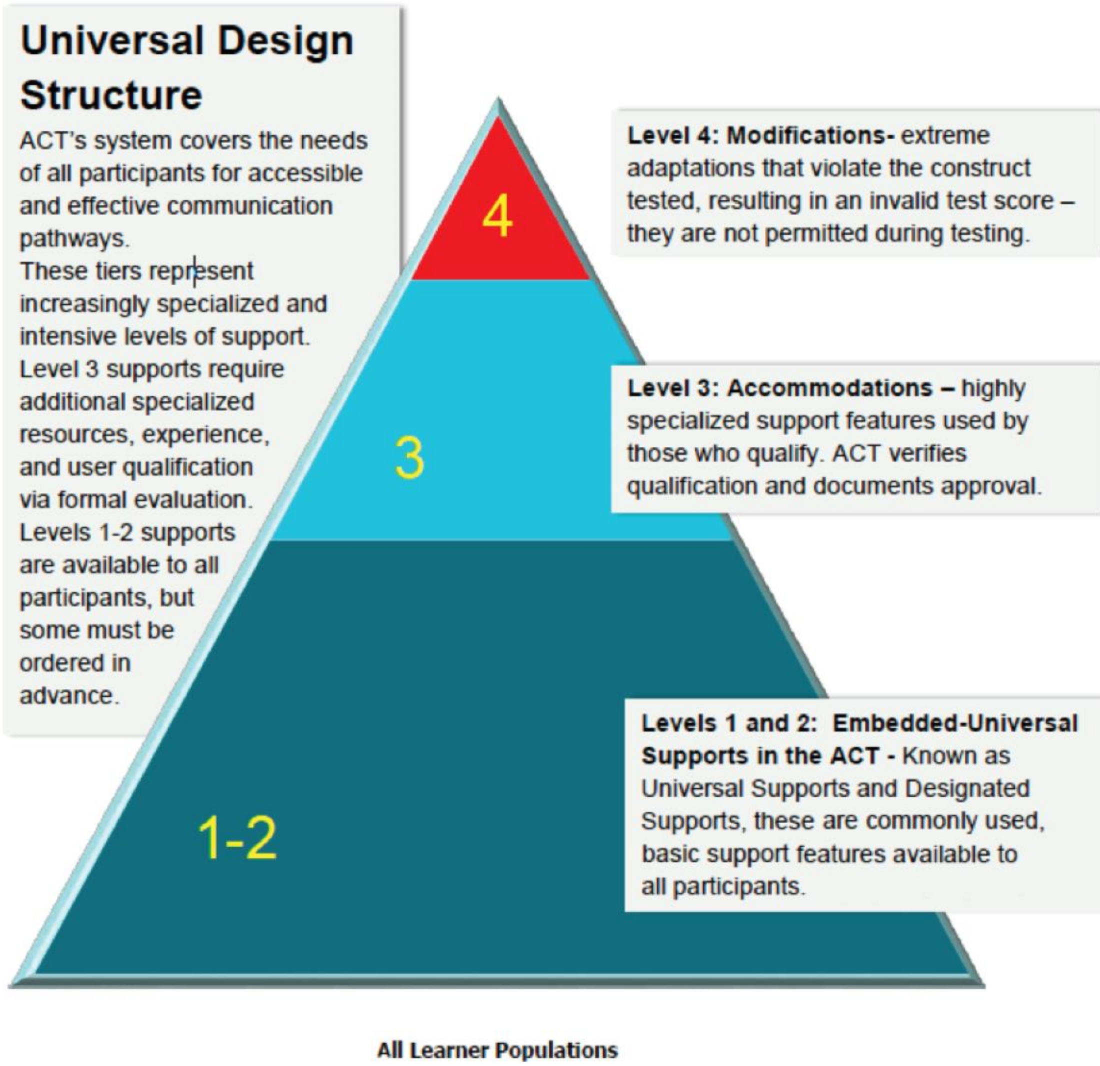

Figure 3.1 The ACT accessibility system structure.

The third level of support, modifications (identified as Level 4 in Figure 3.1), 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 3.1 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)
- color contrast and highlighter 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 they are 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 3.1). 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.

3.1.2 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.

Tables 3.1–3.4 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 3.4). 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.

Table 3.1 Presentation Supports

	Support level		*Content area*					*Applies to:*
Description	*Paper*	*Online*	*Reading*	*English*	*Writing*	*Math*	*Science*	*State & District*
Audio-Recording, Full Test	A	—	—	—	—	—	—	SD
Reader Script, Full Test	A	—	—	—	—	—	—	SD
Screen Reader	A	—	—	—	—	—	—	SD
Text to Speech (available Spring 2018)	—	A	—	—	—	—	—	SD

Description	Support level		Content area					Applies to:
	Paper	Online	Reading	English	Writing	Math	Science	State & District
Translated Written Directions—12 Languages Provided (ELs) [6]	A	A	—	—	—	—	—	SD
Translated Audio, Full Test	*A1*	*A1*	*No*	*No*	*1*	*1*	*1*	*SD*
Word-to-Word Dictionary (ELs) [6]	*A*	*A*	—	—	—	—	—	*SD*
American Sign Language (ASL), Directions Only [7]	*E7*	*A*	—	—	—	—	—	*SD*
Signed Exact English (SEE), Directions Only	*A*	*A*	—	—	—	—	—	*SD*
Signed Exact English (See), Full Test	*A*		—	—	—	—	—	
Cued Speech	*A*		—	—	—	—	—	*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]	*A2*	*A2*	—	—	—	—	—	*SD*
Unified English Braille (UEB), available with Tactile Graphics and Nemeth code for Math and Science (Contracted) [2]	*A*	*A2*	—	—	—	—	—	*SD*
Tactile Graphics (stand-alone) with EBAE/ Nemeth [2]	*A2*	*A2*	—	—	—	—	—	*SD*
Tactile Graphics (stand-alone) with UEB/ Nemeth [2]	*A2*	*A2*	—	—	—	—	—	*SD*
Large Print	*A*		—	—	—	—	—	*SD*
Browser Zoom Magnification		*E*	—	—	—	—	—	*SD*
Magnification	*A*	*E*	—	—	—	—	—	*SD*

Description	Support level		Content area					Applies to:
	Paper	*Online*	*Reading*	*English*	*Writing*	*Math*	*Science*	*State & District*
Line Reader (Straight-edge Tool, locally provided)[7] (Online tool, or Locally provided paper straight edge)	*E7*	*E*	—	—	—	—	—	*SD*
Color Contrast (Online [3]) or Color Overlay (Locally Provided) [7]	*E7*	*3*	*3*	*3*	*3*	*3*	*3*	*SD*

Table 3.2 Interaction and Navigation Supports

Description	*Support level*		*Content area*					*Applies to:*
	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	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 3.3 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*	*No5*	na	na	SD

Table 3.4 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]	*E7*	*E7*	—	—	—	—	—	SD
Special Seating/Grouping [7]	*E7*	*E7*	—	—	—	—	—	SD
Location for Movement [7]	*E7*	*E7*	—	—	—	—	—	SD
Individual Administration [7]	*E7*	*E7*	—	—	—	—	—	SD
Administration at Optimum Time of Day [7]	*E7*	*E7*	—	—	—	—	—	SD
Administration from Home or Care Facility [7]	*E7*	—	—	—	—	—	—	SD
Separate Setting or Location (Familiar Setting and/or Small Group) (ELs) [6, 7]	*E7*	*E7*	—	—	—	—	—	SD
Audio Amplification	*A*	*A*	—	—	—	—	—	SD
Special Lighting [7]	*E7*	*E7*	—	—	—	—	—	SD

	Support Level		Content Area					Applies to:
Description	Paper	Online	Reading	English	Writing	Math	Science	State & District
Adaptive Equipment or Furniture [7]	E7	E7	—	—	—	—	—	SD
Wheelchair Accessible Room	E7	E7	—	—	—	—	—	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 a **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 3.1–3.4:

[1] Provided only as part of the State and District negotiated contract for only nonreportable scores.

[2] All users with blindness will need to use companion paper form braille/tactile graphics on the mathematics and science tests, as critical interpretive information within math and science graphics will not be read aloud. This is required for both the 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 the 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, translating data) are intended to be performed without the use of graphing functionalities often present on calculators.

[5] The ACT writing test domain Language Use & Conventions (including grammar, syntax, and word usage) can be compromised by device usage. Reading, English, mathematics, and science are currently in multiple-choice 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.

3.2 English Learner Supports

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.

3.2.1 Eligibility

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
    - the ability to meet the challenging State academic standards;
    - the ability to successfully achieve in classrooms where the language of instruction is English; or
    - 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 for 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 a half)
- ACT-approved word-to-word bilingual dictionary (no definitions)
- written test instructions translated into written supplements in the student's native language.
  - 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 ELs and receive supports will receive a college-reportable score. ACT Score Reports do not include any specific information about the supports provided.

3.3 Accommodations, Score Validity, and Usage

Since the enactment of the Individuals with Disabilities Education Act (IDEA) in 1975, the total percentage of students with disabilities enrolled in public schools 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.

3.3.1 Differential Prediction for Students Testing with Accommodations

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 the 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 valid FYGPAs and valid ACT Composite scores 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 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.

Results

When jointly using ACT Composite scores and HSGPAs to predict FYGPAs, the mean error of prediction (i.e., 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 test 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 HSGPAs is a good model to predict actual college FYGPAs for both students testing with accommodations and those testing without accommodations. Full results can be found in *Examining the Validity of ACT Composite Score and High School Grade Point Average for Predicting First-Year College GPA of Special-Tested Students*.

Table 3.5. 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 disability						
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)

ACT aims to ensure that all examinees may equally access 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.

References

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

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

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

Snyder, T. D., & Dillow, S. A. (2013). *Digest of education statistics 2012* (NCES 2014–2015). National Center for Education Statistics, Institution of Education Sciences, US Department of Education, Washington, DC.

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

Chapter 4

Test Administration

Chapter 4: Test Administration

4.1 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.

4.2 Administration Windows

For Wisconsin's ACT state testing in the spring, the ACT is administered during predetermined dates/windows. The ACT is administered only on the days/windows 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.

4.3 Modes

In 2019, the students in Wisconsin participating in the state testing administration took the ACT on paper. State testing sites have the option of administering the test on paper or online. The ACT administered online is the same test as the paper version but presented in an online delivery format. Online testing of the ACT is 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.

4.3.1 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.

4.3.2 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.

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 the 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 post-test 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.

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, but that 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.

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.

4.3.3 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.

4.4 Policies and Procedures

4.4.1 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; and
- 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.

4.4.2 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.

4.4.3 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.

4.5 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) the 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.

References

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

Chapter 5

Scaling and Equating

Chapter 5: Scaling and Equating

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

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

5.2.1 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.

5.2.2 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.

5.2.3 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 2. The Composite, STEM, and ELA scores are also comparable across forms because the scores used to compute them have been equated.

5.3 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. Since September 2016, when the 2–12 rounded average domain scores replaced the 1–36 scores for the ACT writing test score reporting, the 1–36 writing scale has solely been used for 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 all prompts administered in the special study were evaluated, 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. For the sum of the domain scores, equating is used to adjust for slight differences in prompt difficulty 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 post-smoothing 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*. Iowa City, IA: ACT.

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: ACT.

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

Chapter 6

Reliability and Measurement Error

Chapter 6: Reliability and Measurement Error

6.1 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.

6.2 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 formula

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

where *k* is the number of test items, s_i^2 is the sample variance of the i[th] item, 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 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.

6.2.1 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 6.1. These values were calculated based on operational test data from the primary and accommodated test forms administered in the 2019–2020 academic year. The reliability estimates are fairly high, with values over 0.9 for English, mathematics, Composite, and ELA scores, and values over 0.8 for reading and science on the primary forms. For the accommodated form, where the distributional characteristics of the students show less variability, the reliability estimates for the overall group are slightly lower, as anticipated.

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

		Primary		Accommodated	
Test	# of items	Reliability	SEM	Reliability	SEM
English	75	0.93	1.65	0.89	1.65
Mathematics	60	0.91	1.59	0.88	1.29
Reading	40	0.88	2.18	0.88	2.04
Science	40	0.86	1.93	0.81	2.03
Composite	215	0.97	0.93	0.96	0.89
ELA	116	0.93	1.41	0.92	1.39

6.2.2 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 based on the 52,614 juniors taking the primary form administered in Wisconsin in the 2019–2020 academic year. For some of the reporting categories, particularly those with very few items, the reliability is low.

Table 6.2. Summary Statistics of Raw Score Reliability and SEM for the ACT Reporting Categories

Test/reporting categories	Number of items	Reliability	SEM
English			
Production of Writing	23	0.80	2.09
Knowledge of Language	12	0.75	1.47
Conventions of Standard English	40	0.88	2.71
Mathematics			
Preparing for Higher Math	36	0.86	2.61
Number & Quantity	6	0.53	1.41
Algebra	8	0.56	1.24
Functions	8	0.63	0.22
Geometry	8	0.59	1.27
Statistics & Probability	6	0.49	1.03
Integrating Essential Skills	24	0.81	2.11
Modeling	21	0.76	1.99
Reading			
Key Ideas & Details	24	0.80	2.18
Craft & Structure	10	0.66	1.39
Integration of Knowledge & Ideas	6	0.59	1.06
Science			
Interpretation of Data	18	0.77	1.82
Scientific Investigation	12	0.65	1.53
Evaluation of Models, Inferences & Experimental Results	10	0.67	1.41

6.2.3 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 6.1 presents the CSEMs for the four multiple-choice tests of the primary forms administered in the 2019–2020 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.

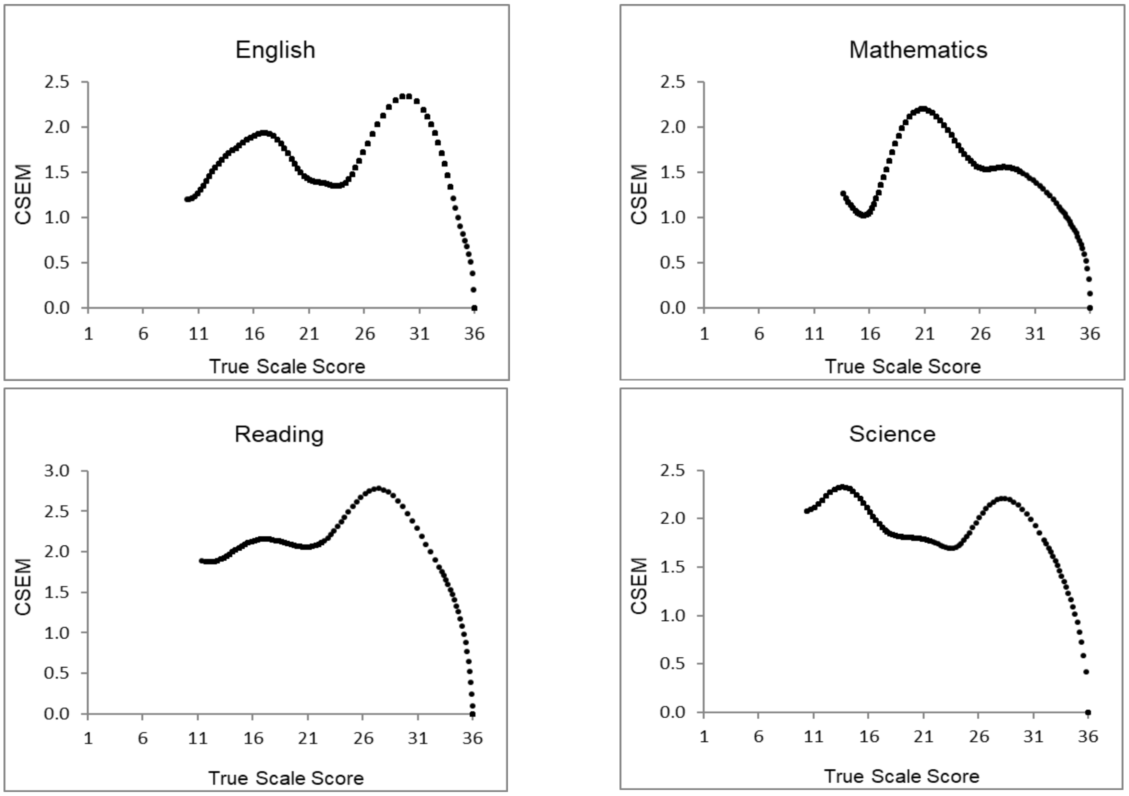

Figure 6.1. CSEM for Multiple-Choice Test Scores

6.2.4 Agreement Indices for the ACT Writing Test Scores

Two major sources can contribute to the measurement error of a writing test score: rater variability and prompt variability. To get a reliability estimate that takes into account both sources of error, a special study is needed where students are administered multiple writing prompts and student responses are rated by multiple raters. Results from such special studies are reported in the ACT Technical Manual (ACT, 2020). With Wisconsin data, where each student takes only a single prompt, only rater agreement is reported based on the students taking the primary form in the 2019–2020 academic year.

As shown in Table 6.3, 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 percentage of students who received the same domain score (from 1 to 6) from both raters, ranged from approximately 0.71 to 0.78 across domains. The perfect plus adjacent agreement rates, or the percentage of students who received either the same domain score or adjacent domain scores (e.g., a score of 5 and a score of 6) from both raters, was as high as 1 for all domains.

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 0.80 to 0.88, indicating good rater agreement.

Table 6.3. Agreement Rates for the ACT Writing Domain Scores

Domain	Agreement index	Value
Ideas & Analysis	Perfect Agreement	0.78
	Perfect + Adjacent Agreement	1.00
	Quadratic Weighted Kappa	0.88
Development & Support	Perfect Agreement	0.78
	Perfect + Adjacent Agreement	1.00
	Quadratic Weighted Kappa	0.86
Organization	Perfect Agreement	0.78
	Perfect + Adjacent Agreement	1.00
	Quadratic Weighted Kappa	0.87
Language Use & Conventions	Perfect Agreement	0.71
	Perfect + Adjacent Agreement	1.00
	Quadratic Weighted Kappa	0.80

6.2.5 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 6.1. The CSEM for the Composite score is plotted as a function of the average of the true scale score 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 6.2. This procedure was repeated for each of the examinees from the primary test forms administered in Wisconsin in the 2019–2020 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 6.2. 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 6.2 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 6.1 for the CSEMs 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 6.2 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 6.2 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.

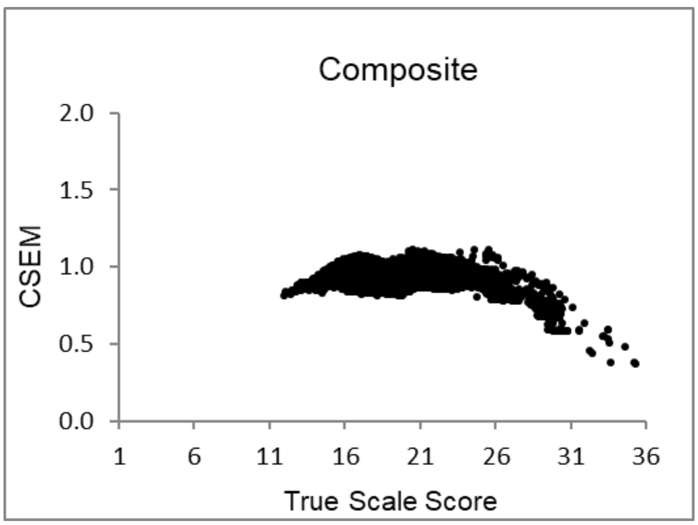

Figure 6.2. CSEM for Composite Scores

6.3 Classification Consistency

Classification consistency refers to the extent to which examinees are classified into the same category over replications of a measurement procedure. Because the same test is 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 the 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.

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

Wisconsin has set their own performance levels in ELA, mathematics, and science for the purpose of score reporting. ACT scores in English, reading, and writing were combined to form the ELA composite score, as shown in Table 6.4. For more information about the performance levels and cut scores, please visit the Wisconsin DPI website (https://dpi.wi.gov/sites/default/files/imce/assessment/pdf/ACT%20Data%20proficiency%20Summary.pdf).

Analyses were conducted to examine the classification consistency on differentiating students into performance levels with the examinees taking the primary test forms administered in Wisconsin in the 2019–2020 academic year. The classification consistencies were calculated using the Livingston and Lewis (1995) method. This method was selected as it can be used in calculating the classification consistency of composite scores, such as the ELA score.

Table 6.5 presents a summary of the agreements between the operational test classifications—that is, the percentages of students who would be consistently classified in the same achievement levels on two equivalent administrations of the test. The agreement rate (percentage consistently classified) was computed for each test score under two classification schemes. One is a two-level classification scheme, which refers to proficient/non-proficient decisions (i.e., Basic and below vs. Proficient and above), and the other is a four-level classification scheme, which refers to classification using all four performance levels (i.e., Advanced, Proficient, Basic, and Below Basic).

Table 6.4. ACT Performance Level Cut Scores for Wisconsin

Test	Basic Cut Score	Proficient Cut Score	Advanced Cut Score
ELA	15	20	28
Mathematics	17	22	28
Science	18	23	28

Table 6.5. Classification Consistency for the ACT Performance Level

		Classification Consistency	
Test	Number of items	Two-level	Four-level
ELA	116	0.91	0.70
Mathematics	60	0.86	0.64
Science	40	0.89	0.76

Classification consistency for the ACT Readiness Ranges was computed for each of the ACT test reporting categories. These values, provided in Table 6.6, are based on data from the 2019–2020 school year.

6.3.2 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). For the primary forms administered in the 2019–2020 academic year, the classification consistency was 0.73, which was moderately high considering the number of items that contribute to UCT scores and the number of performance levels. There are 23 UCT items in the primary form, and the percentages of students classified as Below Proficient, Proficient, and Above Proficient were 53%, 28%, and 18%, respectively.

6.3.3 Classification Consistency for Progress Toward ACT NCRC Indicator

The second indicator, Progress Toward the ACT National Career Readiness Certificate (ACT NCRC), had a classification consistency value of 0.79 on the primary forms administered in the 2019–2020 academic year. This value is quite high given that there are four performance levels for the ACT NCRC, as shown in Table 6.7. Note that the classification consistency index is an indication of the stability of the Progress Toward the 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 6.6. Classification Consistency for the ACT Readiness Ranges

Test/Reporting Categories	Number of items	Classification Consistency
English		
Production of Writing	23	0.81
Knowledge of Language	12	0.79
Conventions of Standard English	40	0.86
Mathematics		
Preparing for Higher Math	36	0.87
Number & Quantity	6	0.70
Algebra	8	0.73
Functions	8	0.76
Geometry	8	0.71
Statistics & Probability	6	0.68
Integrating Essential Skills	24	0.82
Modeling	21	0.80
Reading		
Key Ideas & Details	24	0.82
Craft & Structure	10	0.75
Integration of Knowledge & Ideas	6	0.72
Science		
Interpretation of Data	18	0.79
Scientific Investigation	12	0.77
Evaluation of Models, Inferences & Experimental Results	10	0.76

Table 6.7. 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

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

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* (3rd ed.). 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 7

Validity Evidence

Chapter 7: Validity Evidence

7.1 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.

ACT scores are comparable across National and State and District administrations, and the reliability and validity information described in Chapters 6 and 7 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.

7.2 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 coursework, 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.

7.2.1 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 Chapter 2, 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.

7.2.2 Statistical Relationships Between ACT Scores and High School Coursework 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.

[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 (ACT, 1989).

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 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, are currently taking, plan to take, or do not plan to take the course. If they have taken the course, they indicate the grade they received (A–F). Self-reported coursework and grades collected with the CGIS have been found to be accurate relative to information provided on student transcripts (Sanchez & Buddin, 2016; Sawyer, Laing, & Houston, 1988; Valiga, 1986; see also the next section).

Table 7.1 displays the ACT scale score means and standard deviations in the English, mathematics, reading, and science tests for three groups of students by the number of years of English, mathematics, social studies, and science coursework the students 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 three and a half years of English and those who expected to take two 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 7.1. Means and Standard Deviations for ACT Scores: 2016 ACT-Tested High School Graduates by Years of Subject-Relevant Coursework

Years of coursework	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 7.2, students who have completed or plan to complete a core curriculum tend to achieve higher ACT scores than those who have not completed a core curriculum (ACT, 2016b), where a core curriculum is defined as 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 three 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 coursework and ACT English and mathematics scores.

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

Academic preparation	Reference year	N	ACT scores				
			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 7.1 and 7.2 support the notion that the ACT is a curriculum-based test. Additionally, an analysis by McNeish, Radunzel, and Sanchez (2015) showed that, in general, coursework and high school grades were strongly associated with performance on the ACT, after statistically controlling for other factors. However, it is also conceivable that some other factors, including noncognitive factors, account for the observed association between high school coursework and ACT scores. In the McNeish et al. study, the researchers investigated the relationships between noncognitive characteristics, high school coursework and grades, school characteristics, and test scores of ACT-tested students. The remainder 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. Twelve percent of the initial sample responded and met the study inclusion criteria: 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.

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 coursework 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 7.3 for the various block groupings denoted in bold font; results for gender and race/ethnicity are shown in Table 7.14). 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, coursework taken, advanced coursework taken, school characteristics, noncognitive characteristics, demographics related to socioeconomic status (SES), 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 7.3. 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 7.1). 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 coursework, such as accelerated, advanced, or honors courses, or courses for college credit, accounted for an additional 3% to 5% of the variance in ACT scores. HSGPA and coursework 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 7.3 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 7.3, 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 points, 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 points 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 coursework in English were expected to score 1.0 to 1.1 points higher on the ACT reading and English tests. In contrast, taking advanced coursework 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, coursework taken, school characteristics, noncognitive characteristics, and demographic characteristics. High school academic factors, such as HSGPA and coursework, 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 was associated with sizable mean ACT score differences. Specific English and social studies courses were not included in the models because of the limited variability in students' course-taking in these subject areas and their collinearity with other variables, such as coursework taken in mathematics and science. The findings from this study are consistent with earlier studies (Noble, Davenport, Schiel, & Pommerich, 1999a, 1999b; Noble & McNabb, 1989; Schiel, Pommerich, & Noble, 1996) that examined coursework, grades, and ACT score relationships.

Table 7.3. 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[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. Math, 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 coursework					
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
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
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*

	ACT score				
Predictor	English	Mathematics	Reading	Science	Composite
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
Need help with educational/occupational plans (yes/no)	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	0.56	0.60	0.44	0.49	0.61
Mean square error	4.22	3.21	4.47	3.54	3.13

Note. Regression coefficients for all achievement, school characteristic, and noncognitive variables were statistically significant ($p < .01$) unless denoted otherwise. Regression coefficients for gender and race/ethnicity are shown in Table 7.14.
† 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; Other 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 coursework includes any accelerated, advanced, honors, and dual-enrollment courses taken in the subject area by the student while in high school.

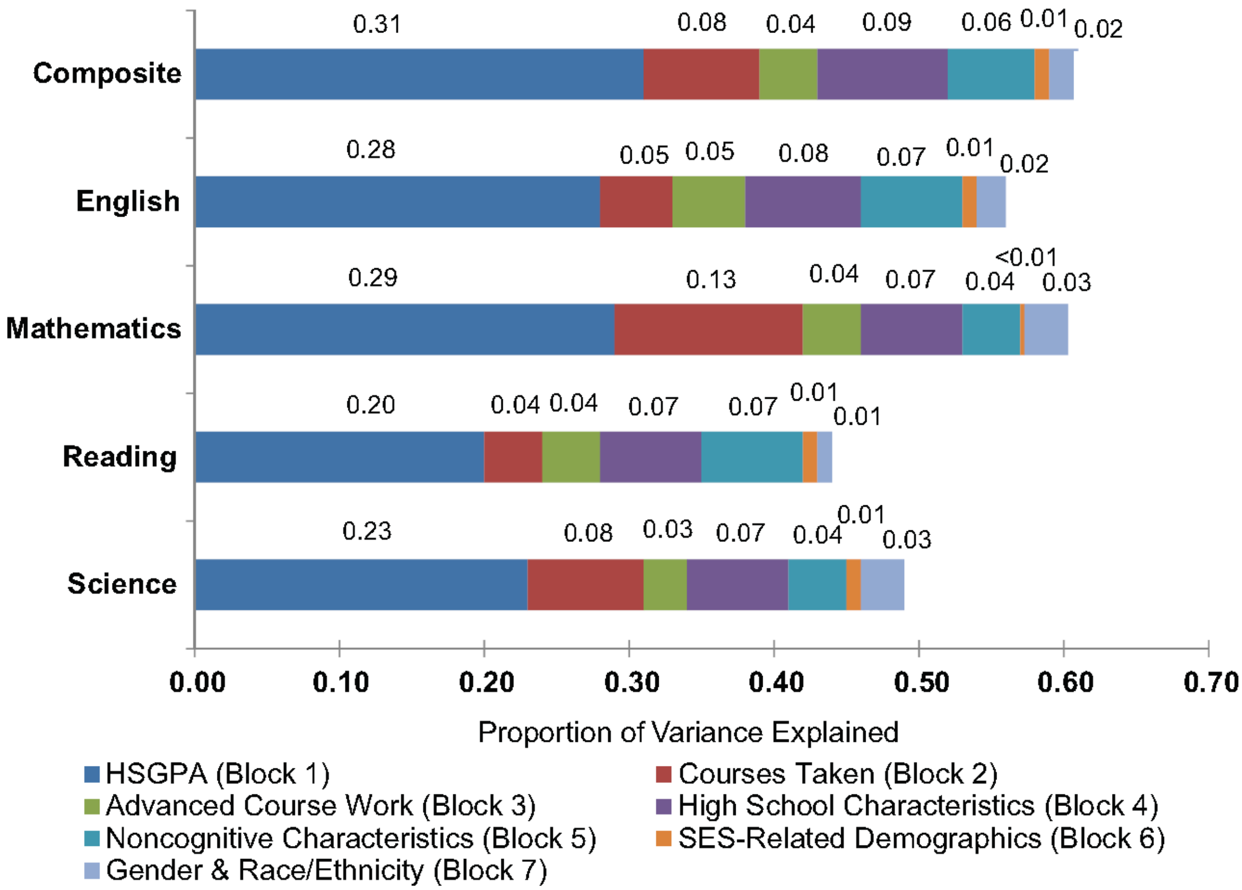

Figure 7.1. Proportion of Variance in ACT Scores Associated With HSGPA, High School Coursework Taken, School Characteristics, Noncognitive Characteristics, and Demographic Characteristics (McNeish et al., 2015)

7.2.3 Construct Contamination in HSGPA

ACT scores are statistically associated with high school grades (Table 7.4; 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 7.4. 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 coursework and grades differed little by gender, race/ethnicity, and income. The results indicated that self-reported coursework and grades are reasonably accurate measures for use in education research and for preliminary screening by college admission officials.

Grade Inflation

Grade inflation is present when grades increase over time without a concomitant increase in 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 remainder 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 7.5 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 7.2 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 studies further, Figure 7.3 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 and District 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 7.5. 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

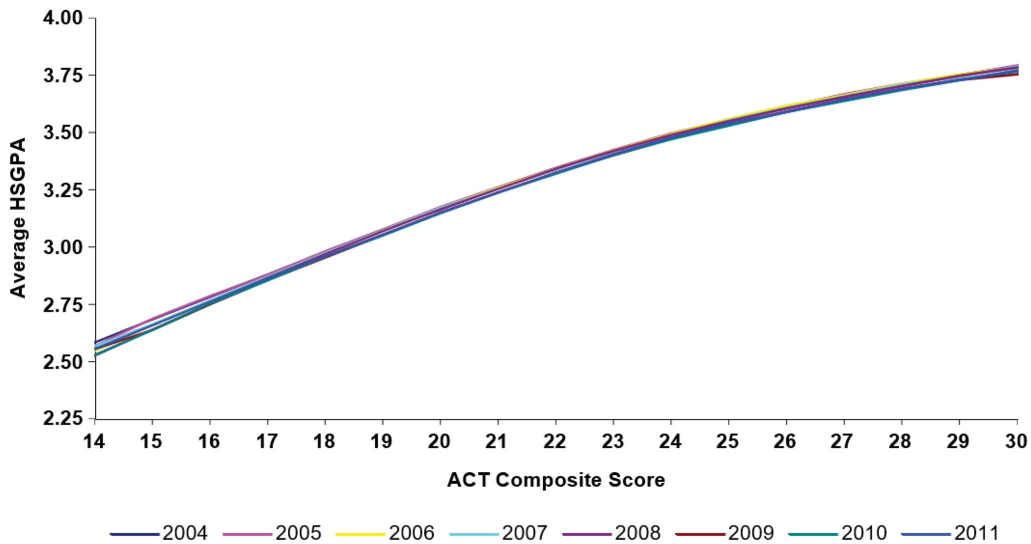

Figure 7.2. Plot of Conditional HSGPA by ACT Composite Score for the Years 2004 to 2011

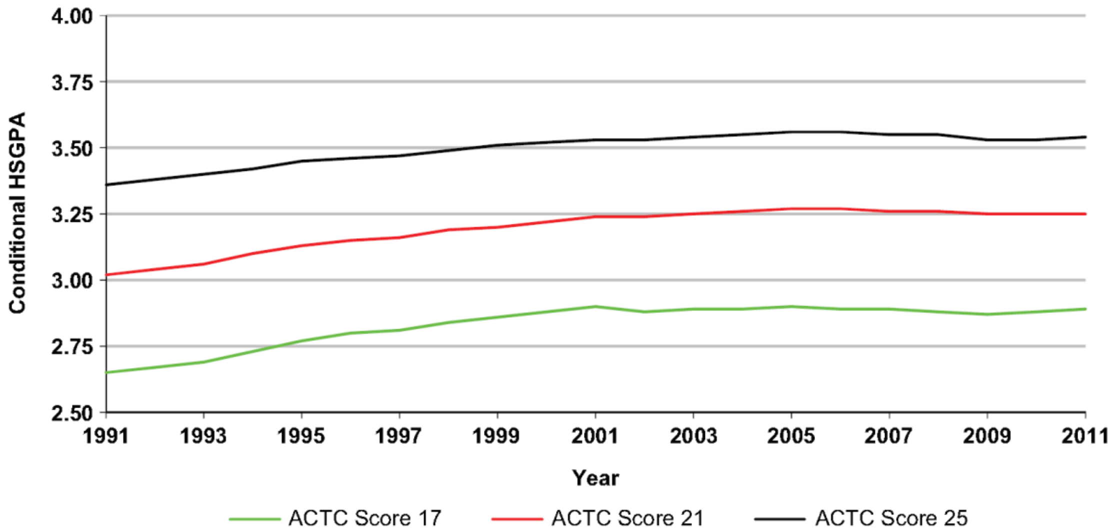

Figure 7.3. Plot of Conditional HSGPA Between 1991 and 2011 for Selected ACT Composite Scores

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 first quintile group (bottom 20% of schools) and those that remained in the fifth quintile group (top 20% of schools) on the ACT Composite score for all five years. The number of schools in the first quintile group and the fifth 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 first quintile group used more lenient grading standards than the schools in the fifth quintile group. HSGPA was regressed on ACT Composite score within each quintile group for each year. If the schools in the first and fifth 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 7.6 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 first quintile group's 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 fifth quintile group.

Figure 7.4 displays the regression lines estimating the linear relationship between overall HSGPA and ACT Composite score in 2000 for the first and fifth quintile groups (denoted Q1 and Q5, respectively). From the figure, it is clear that for students with the same ACT Composite score, the first quintile group had a higher mean overall HSGPA than the fifth 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 7.6. 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	86536	.59	0.073	1.56	

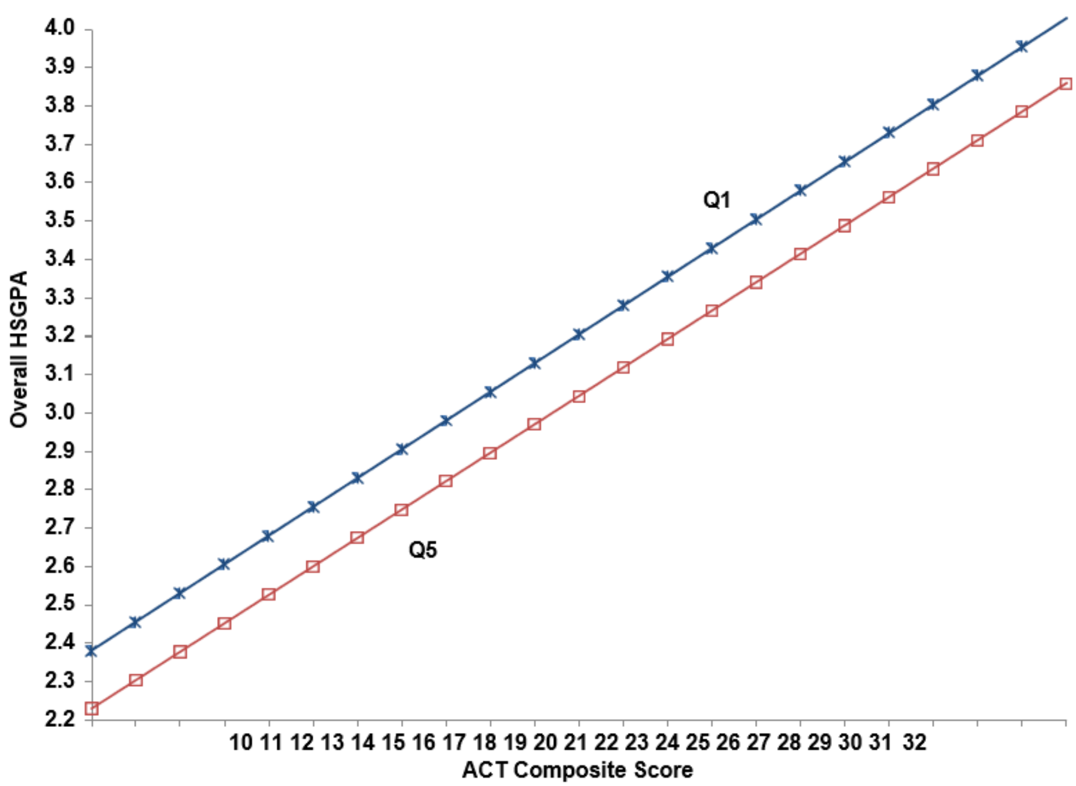

Figure 7.4. Plot of the Year 2000 Linear Regressions of Overall HSGPA on ACT Composite Score

7.2.4 Statistical Relationships Between ACT Benchmark Attainment and High School Coursework 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 9.

A study by Ling and Radunzel (2017) examined how the high school coursework taken and grades earned related to students' chances of meeting the ACT College Readiness Benchmarks in each of the four 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 7.2.2 for more details about the study sample).

At the time they registered to take the ACT, students provided other information, such as high school coursework taken and grades earned. Students' readiness for college coursework 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 coursework 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 coursework 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 7.7). 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 coursework was primarily associated with meeting the ACT Benchmark in mathematics (Table 7.7). 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 coursework 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 coursework taken in high school accounted for between 7% (reading) and 16% (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 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 7.7. 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)	1.38	1.70	1.19**	1.47
Algebra 1, Algebra 2, Geometry & other Advanced Math	1.28**	1.69	1.14**	1.42
Algebra 1, Algebra 2, Geometry & Trig	2.07	3.09	1.58	2.21
Algebra 1, Algebra 2, Geometry, Trig & other Advanced Math	1.73	4.52	2.05	3.02
Algebra 1, Algebra 2, Geometry, Trig & Calculus	1.92	5.00	2.07	3.34
Algebra 1, Algebra 2, Geometry, other Advanced Math, Trig & Calculus	1.38	1.70	1.19**	1.47
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 coursework[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 coursework includes any accelerated, advanced, honors, and dual-enrollment courses taken in the subject area by the student while in high school.

7.2.5 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 coursework taken in high school for increasing ACT performance depend on which 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 used only 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 7.8). 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 7.9), 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 7.10). 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 7.11). 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 7.2.2), 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 coursework and grades.

Table 7.8. Predicting ACT English Scores

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 7.9. Predicting ACT Mathematics Scores

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 7.10. Predicting ACT Reading *Scores*

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 7.11. Predicting ACT Science *Scores*

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

7.2.6 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 coursework, course grades, student and high school characteristics, and educational plans by race/ethnicity, gender, and/or annual family income (e.g., Noble et al., 1999a, 1999b; Noble, Crouse, Sawyer, & Gillespie, 1992; Noble & McNabb, 1989; Chambers, 1988). Their findings suggest that differential performance may be largely attributable to differential academic preparation across student demographic groups.

Table 7.12 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 7.12. Descriptive Statistics for ACT Composite Scores by Racial/Ethnic Group, 2015–2016

			Average Composite score	
Ethnic group	% with core or more[a]	% with HSGPA ≥3.0[b]	Core or more	Less than core
African American/Black	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

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

The extent to which ACT scores vary by gender has also been examined (Table 7.13). 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 the differential performance of various subgroups is potentially explained by factors such as academic achievement, courses taken, school characteristics, and noncognitive variables.

Table 7.13. Average ACT Scores by Gender, 2012–2016

Gender	Reference year	*N*	ACT score				
			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 7.2.2 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 coursework 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 7.1; 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 coursework taken and HSGPA. As illustrated earlier in Figure 7.1, 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 7.14; 1% for the SES-related demographic characteristics and 1% to 3% for gender and race/ethnicity combined).

Table 7.14 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 7.3. 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 7.14. 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 R2 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 R2 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 the cognitive, school-level, and noncognitive variables shown in Table 7.3.
† 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 were <$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 7.5). 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).

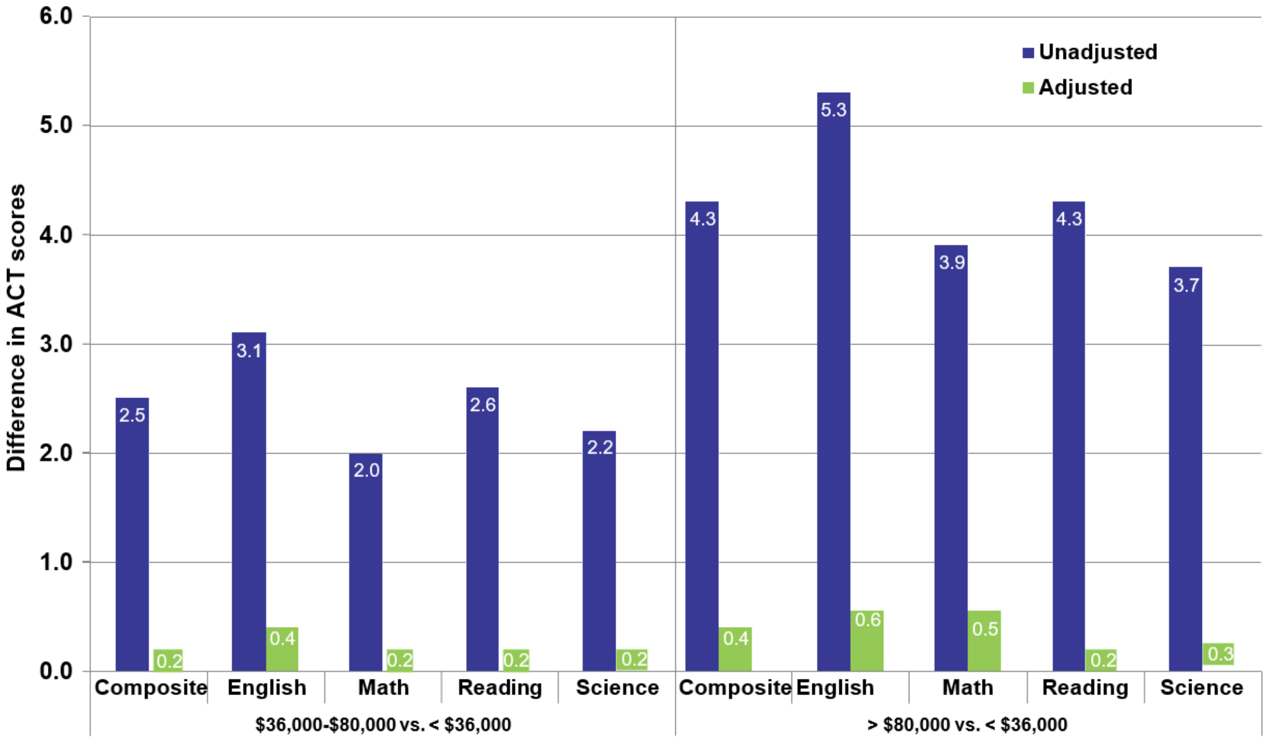

Figure 7.5 Unadjusted and adjusted mean differences in ACT scores by family income

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).

7.3 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.

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.

7.3.1 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 6). As an example, if the observed correlation between the ACT Composite score and FYGPA is .38, the reliability estimate for FYGPA is .81, and the reliability of ACT scores is set to 1.0 (no correction), the validity coefficient for ACT Composite scores would increase from .38 to .42 (.38/(√(.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 7.15 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 7.15. 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 7.16 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 7.16. 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

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

where *W* = 1, if a student is successful in college
W = 0, if a student is not successful in college, and
X is the student's admission test score.

An example of an estimated logistic function is the curve labeled "Probability of C or higher" in Figure 7.6. 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 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

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

where $\hat{P}[W = 1|X = x]$ 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 $\hat{P}[W = 1|X = x]$ 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.

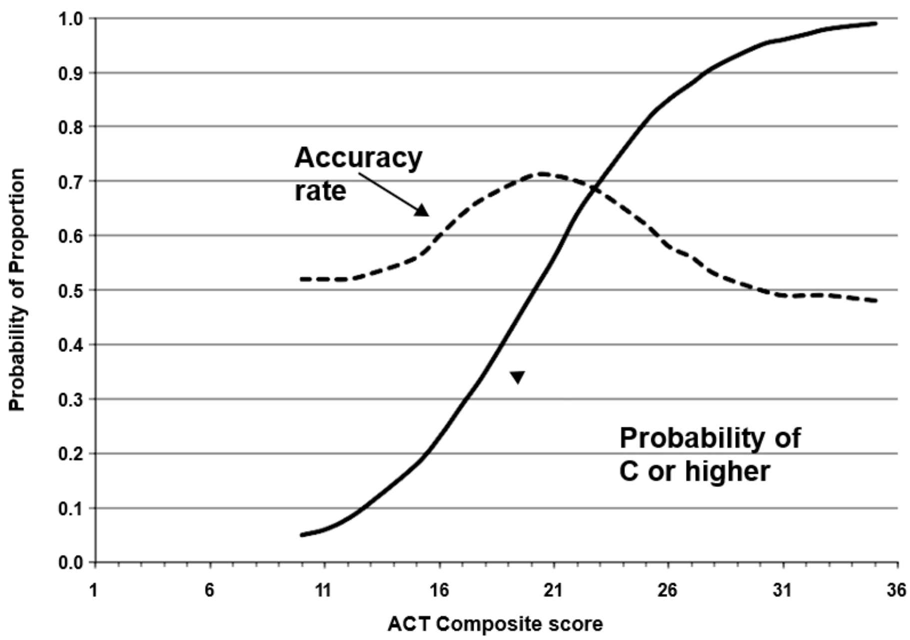

Figure 7.6. Probability of C or higher FYGPA and accuracy rate

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 7.6 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 non-enrolled 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 7.11 and 7.12 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.

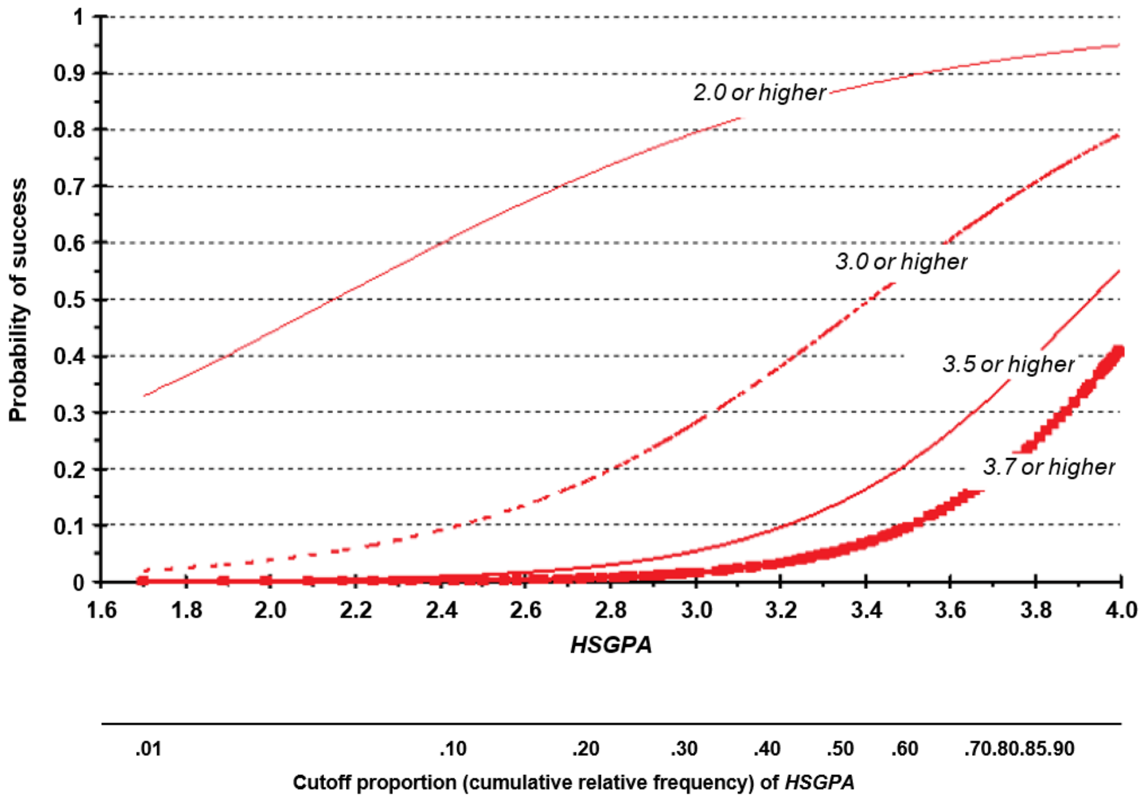

Figure 7.7. Probabilities of success associated with 2.0, 3.0, 3.5, and 3.7 or higher FYGPA and being retained through the first year, based on HSGPA

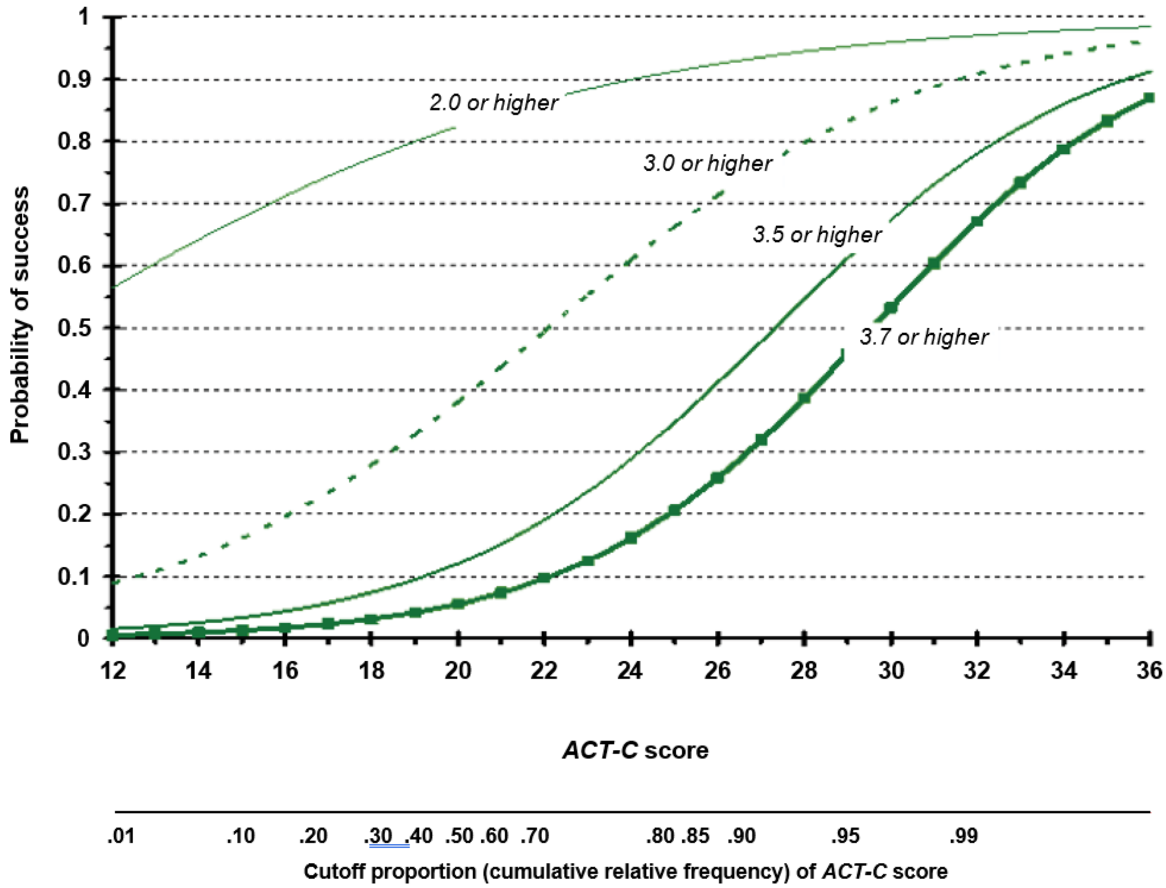

Figure 7.8. Probabilities of success associated with 2.0, 3.0, 3.5, and 3.7 or higher FYGPA and being retained through the first year, based on ACT Composite score

Figure 7.9 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.

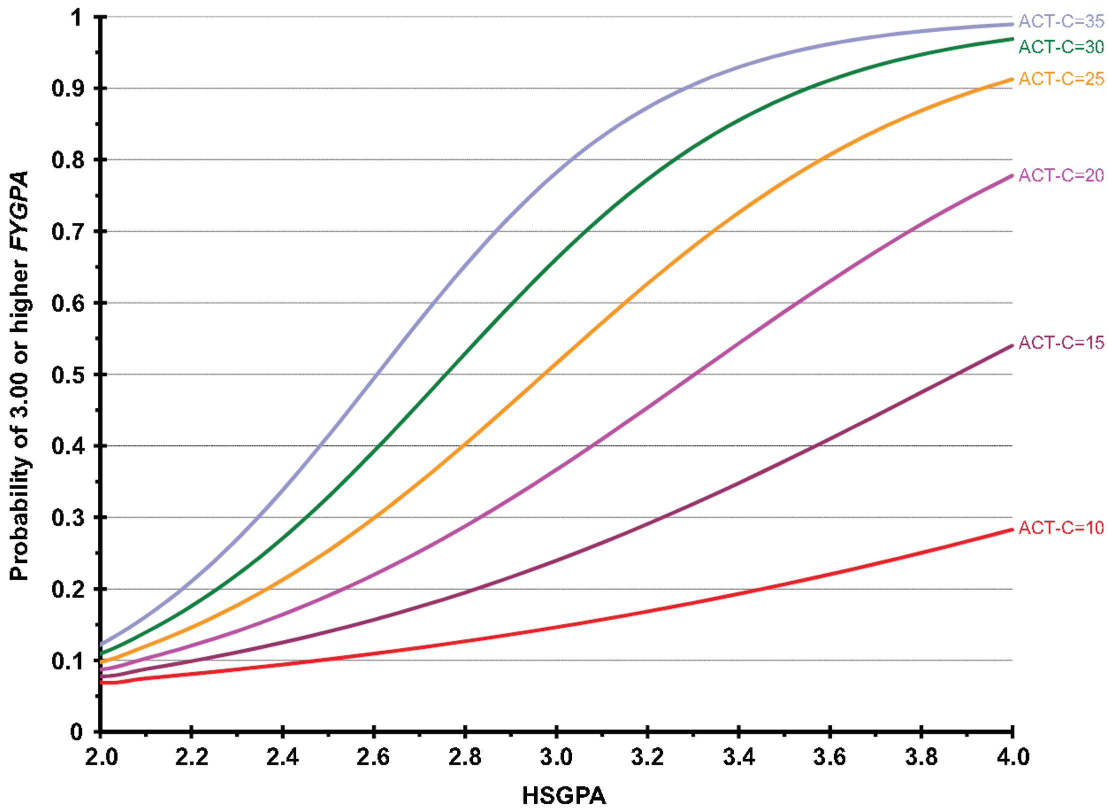

Figure 7.9. Probabilities of success associated with 3.0 or higher FYGPA and being retained through the first year, based on HSGPA and ACT Composite score

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 .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 7.17 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 7.18 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 7.17. Median Incremental Success Rate with Respect to Base Success Rate, by FYGPA Success Level, Cutoff Proportion, and Selection Variable (N = 192)

	Approx. value of		Success Level											
			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 7.18. 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)

	Approx. value of		Success Level											
			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	.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	

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).

7.3.2 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 7.3.1 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 7.14 and 7.15). 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 under-prediction 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 7.11). 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 under-prediction for White students and over-prediction for both Hispanic and African American students (see Table 7.19). Across institutions, for the 2.5 or higher success level, Hispanic students showed the least amount of over-prediction. African American students, however, showed evidence of moderate over-prediction. For the 3.0 or higher success level, the over-prediction 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 7.19). 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.

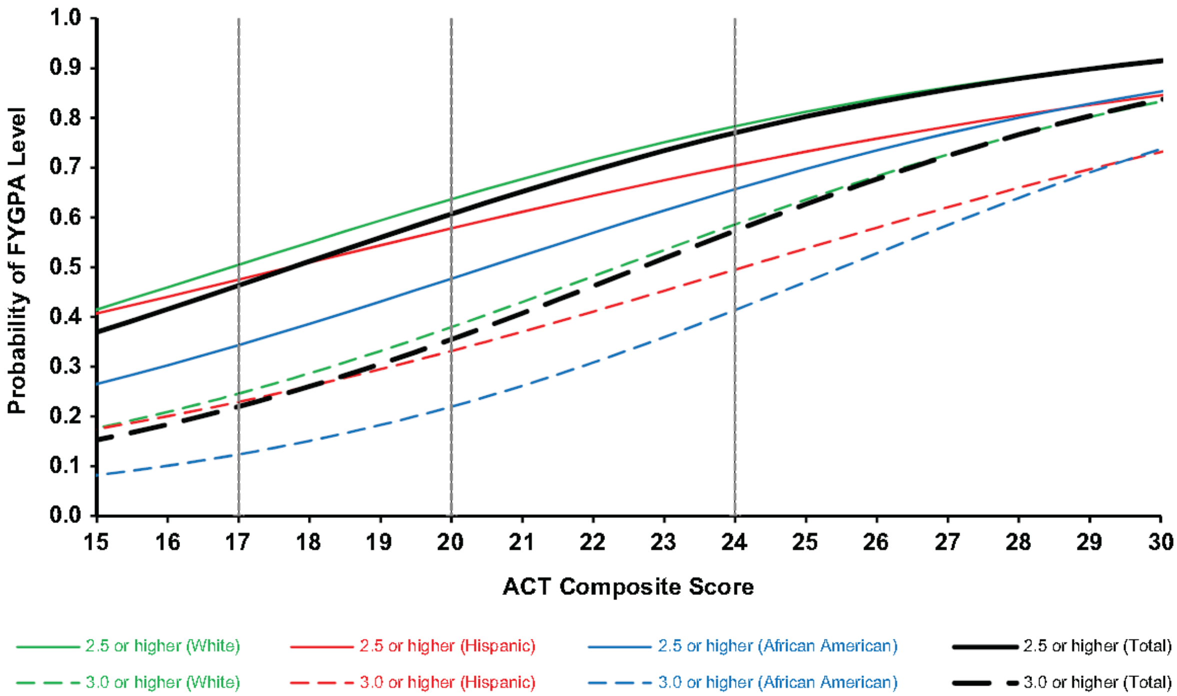

Figure 7.10*.* Estimated probabilities of achieving specific FYGPA levels based on ACT-C score, by race/ethnicity

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

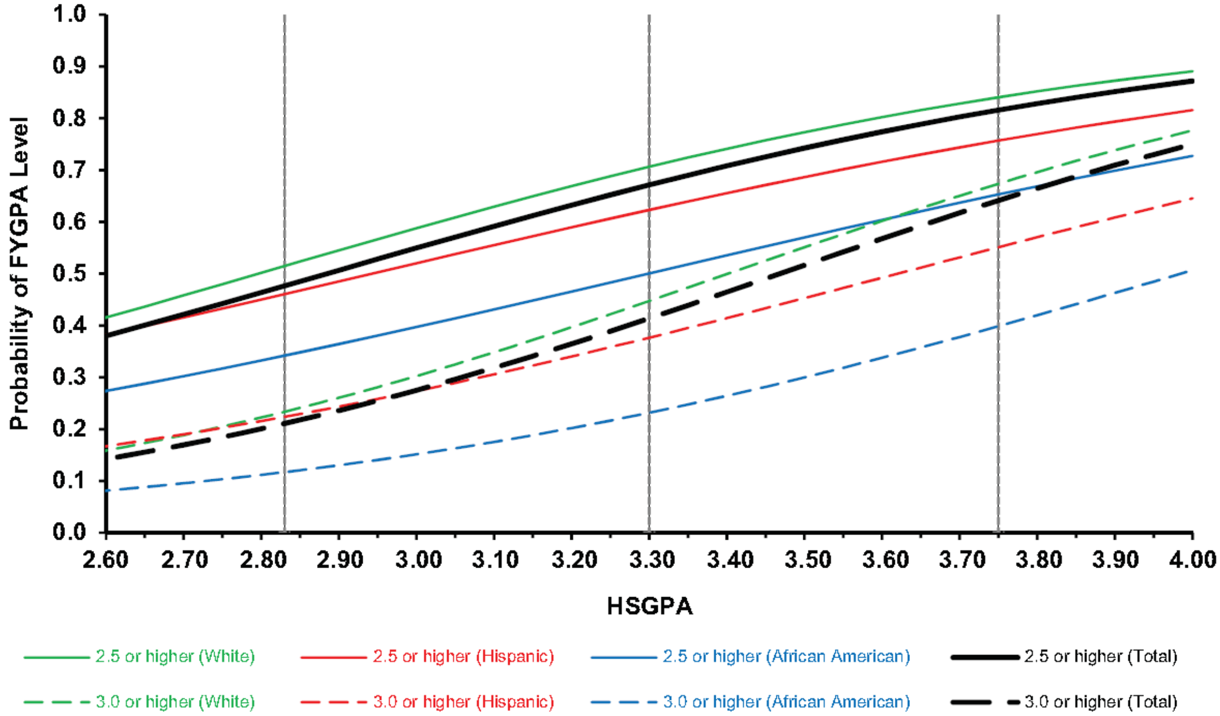

Figure 7.11. Estimated probabilities of achieving specific FYGPA levels based on HSGPA, by race/ethnicity

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

Table 7.19. Median Statistics for Predicting Specific Levels of FYGPA by Race/Ethnicity across Institution

				Subgroup-specific probability 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)
2.5 or higher FYGPA								
ACT-C	242	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		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)

Predictor variable	N	Total-group cutoff	Race/ ethnicity	Subgroup-specific probability 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)
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)
3.0 or higher FYGPA								
ACT-C	236	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		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)

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 over-prediction for males than under-prediction for females. As shown in Table 7.20, using a total-group cutoff score under-predicted the probability of success for females and over-predicted the probability of success for males for both success levels. Across institutions, the use of ACT-C alone resulted in a slightly larger differential prediction than when HSGPA was used in isolation.

Table 7.20. Median Statistics for Predicting Specific Levels of FYGPA by Gender across Institutions

				Subgroup-specific probability success	Maximum accuracy rate (AR)	Increase in AR (ΔAR)	Success rate (SR)	Observed percentage below OC (PB)
Predictor variable	N	Total-group cutoff	Gender	Median (Min/Max)	Median (Min/Max)	Median (Min/Max)	Median (Min/Max)	Median (Min/Max)
2.5 or higher FYGPA								
ACT-C	253	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		2.8	Female	0.53 (0.16/0.61)	73 (57/96)	6 (0/51)	75 (53/96)	24 (0/89)
			Male	0.47 (0.18/0.59)	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	247	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		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)

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 under-predicted, respectively (see Table 7.21). Relatively little evidence of over- or under-prediction 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 7.21. Median Statistics for Predicting Specific Levels of FYGPA by Income across Institution

				Subgroup-specific probability success	Maximum accuracy rate (AR)	Increase in AR (ΔAR)	Success rate (SR)	Observed percentage below OC (PB)
Predictor variable	N	Total-group cutoff	Income	Median (Min/Max)	Median (Min/Max)	Median (Min/Max)	Median (Min/Max)	Median (Min/Max)
2.5 or higher FYGPA								
ACT-C	253	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		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)
3.0 or higher FYGPA								
ACT-C	247	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		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	0.5 (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 over-predicts the probability of success of Hispanic and African-American students, males, and lower-income students. Both ACT-C and HSGPA slightly under-predict 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 Validity

Evidence for examinees with testing accommodations is described in Chapter 3. Briefly, Huh and Huang (2016) found that ACT tests scores obtained under accommodations for students with disabilities are predictive of FYGPA. Furthermore, a prediction model containing both ACT Composite scores and HSGPA is a good model to predict actual college FYGPA for both students testing with accommodations as well as those testing without accommodations.

7.4 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.

7.4.1 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.

7.4.2 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 7.3.1 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 7.22 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) were 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.

The optimal cutoff score for a given course varies across institutions (Allen, 2013). Variation in grading standards and course difficulty across institutions can contribute to this variation in optimal cut scores. Because results vary across institutions, institutions should collect their own course outcome data and determine their placement cutoff scores accordingly. For more details on methods for setting institution-specific cut scores, see section 7.4.5.

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.

Table 7.22. 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	Q1	Med.	Q3	Q1	Med.	Q3
English courses													
Composition I	English	215	18	0.135	63	67	72	1	2	9	63	68	73
Composition I		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.

7.4.3 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 7.16 and 7.17 illustrate the value of using multiple measures when estimating a student's likelihood of course success. Figure 7.12 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 7.13, 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).

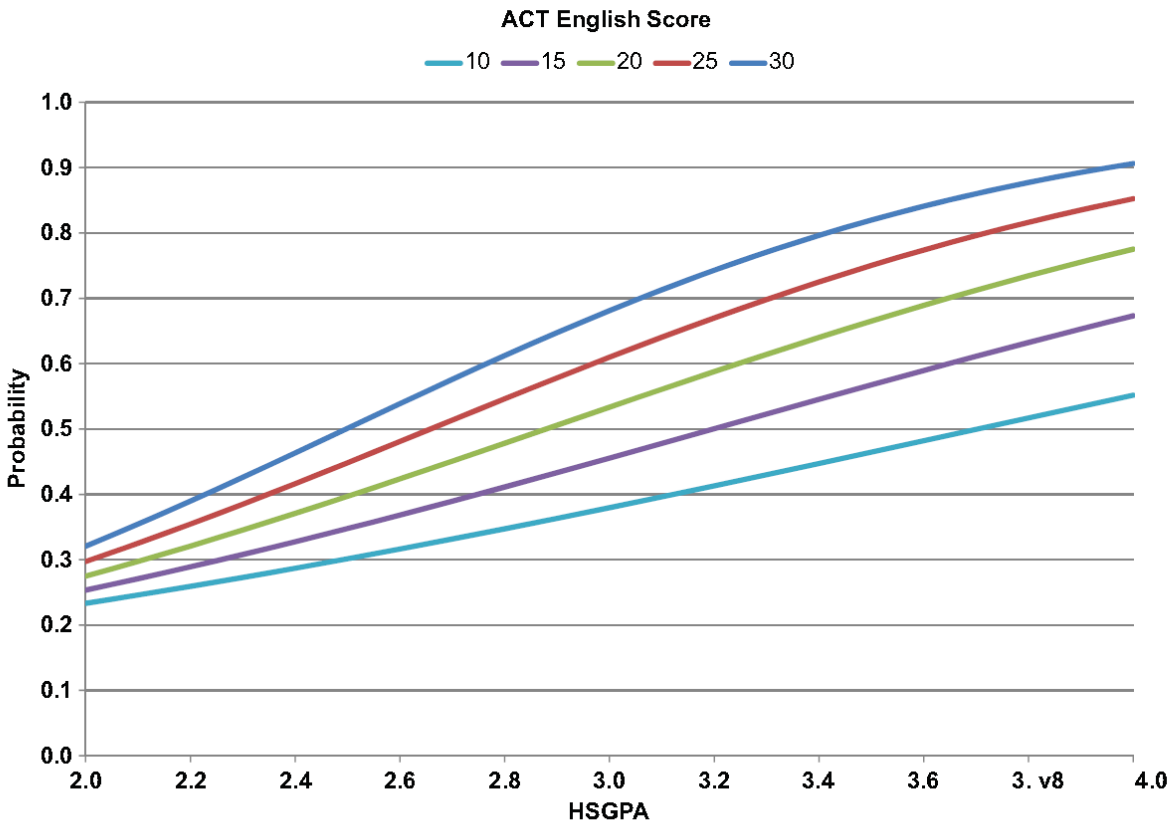

Figure 7.12. Probability of earning a grade of B or higher in English Composition I at two-year institutions, given ACT English score and HSGPA

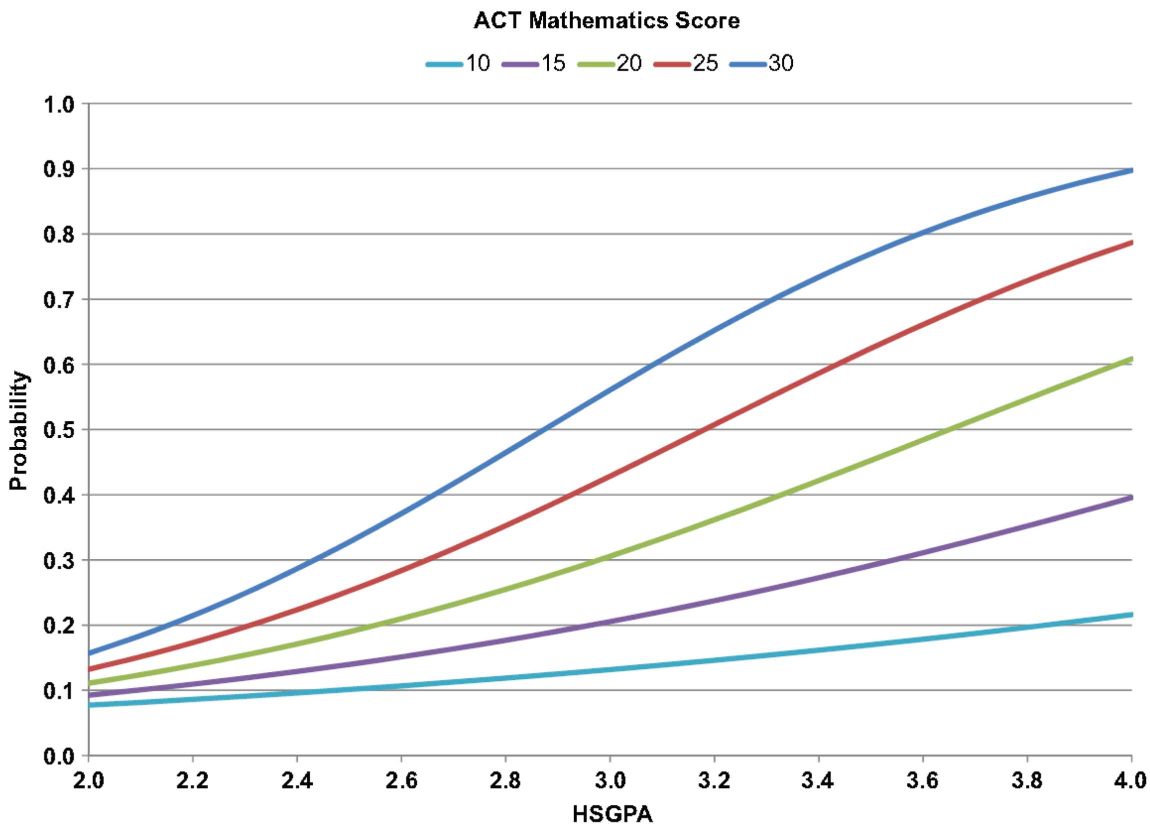

Figure 7.13. Probability of earning a grade of B or higher in College Algebra at two-year institutions, given ACT mathematics score and HSGPA

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 7.23. 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 7.23. 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	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

7.4.4 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 7.14 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.

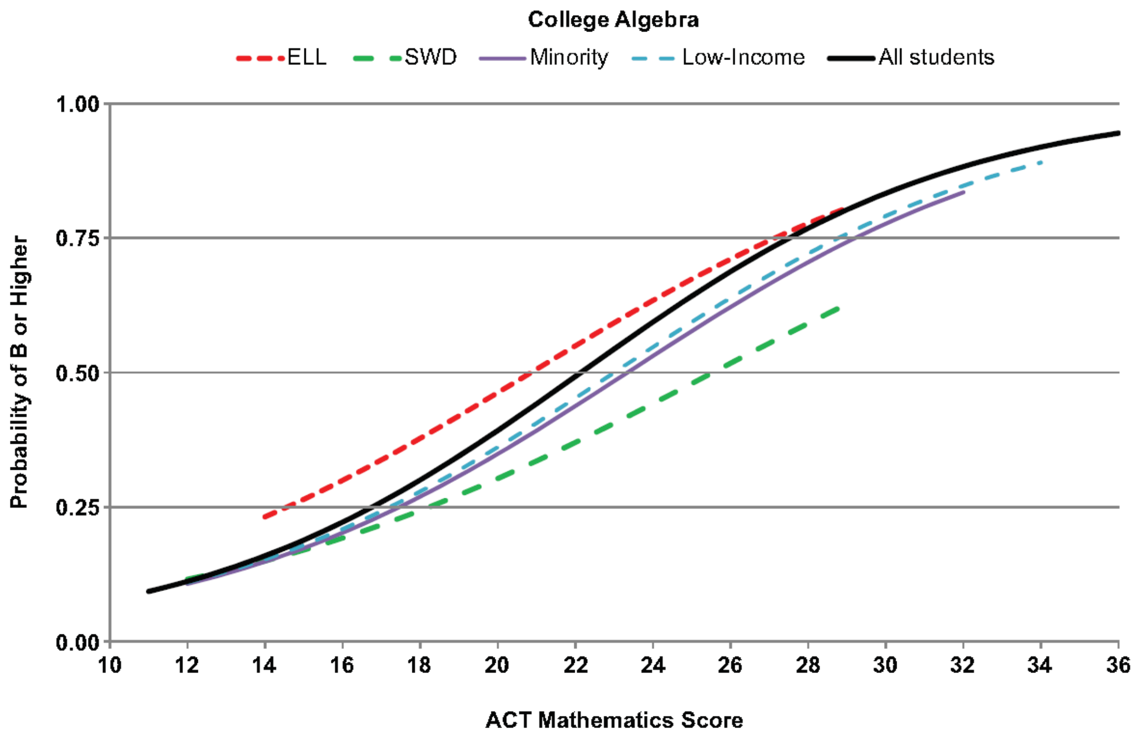

Figure 7.14. Probability of earning a grade of B or higher in College Algebra by ACT mathematics score and student demographic group (ELL is for English language learners; SWD is for students with disabilities)

Table 7.24 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, over-prediction 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, under-prediction occurs when the 50% success cut score for a demographic group is lower than the 50% cut score for the total group.

Table 7.24. 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 under-prediction for English language learners (Mattern, Patterson, Shaw, Kobrin, & Barbuti, 2008; Patterson & Mattern, 2012) and slight over-prediction 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-/under-prediction 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).

7.5 Evaluating Students’ Likelihood of College Success

Sections 7.3 and 7.4 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.

7.5.1 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 9 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 7.15; by 23 to 29 percentage points); (b) earn a B or higher grade in first-year college courses (Figure 7.16; by 18 to 27 percentage points); (c) achieve a FYGPA of 3.0 or higher (Figure 7.17; by 23 to 27 percentage points), and (d) remain enrolled at the same institution in year two (Figure 7.18; 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 7.25). 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.

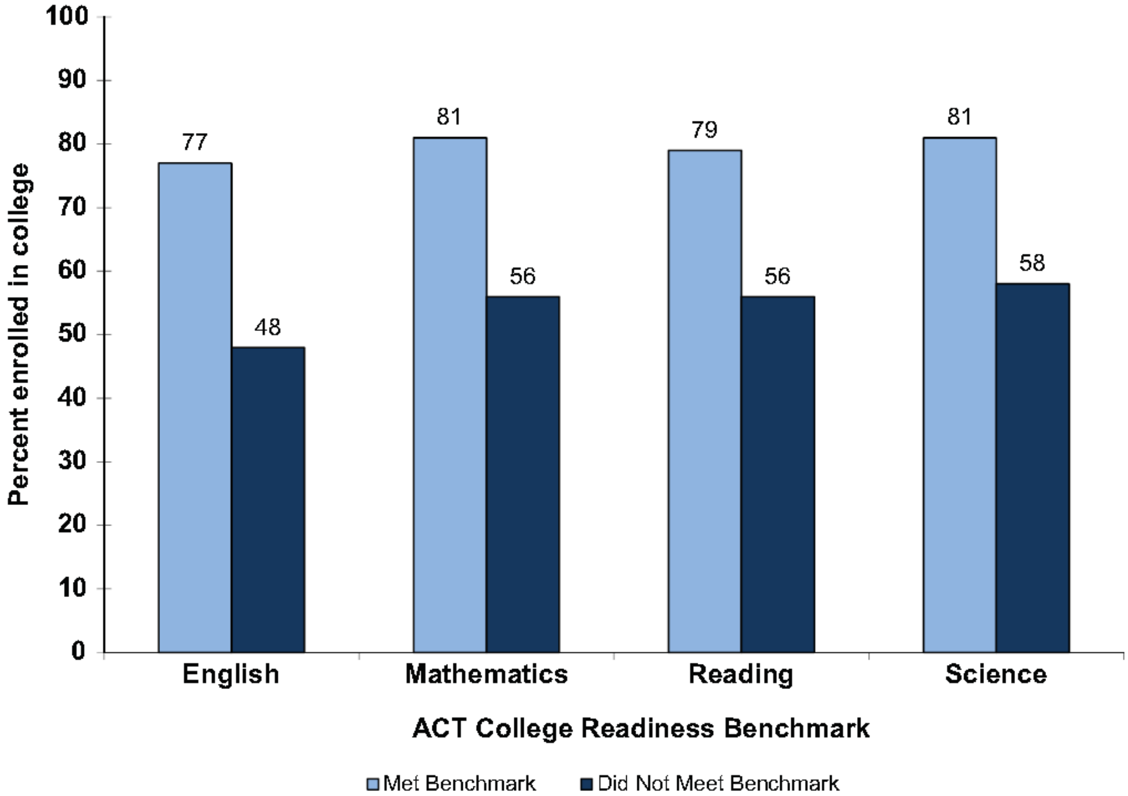

Figure 7.15. College enrollment rates by ACT College Readiness Benchmark attainment

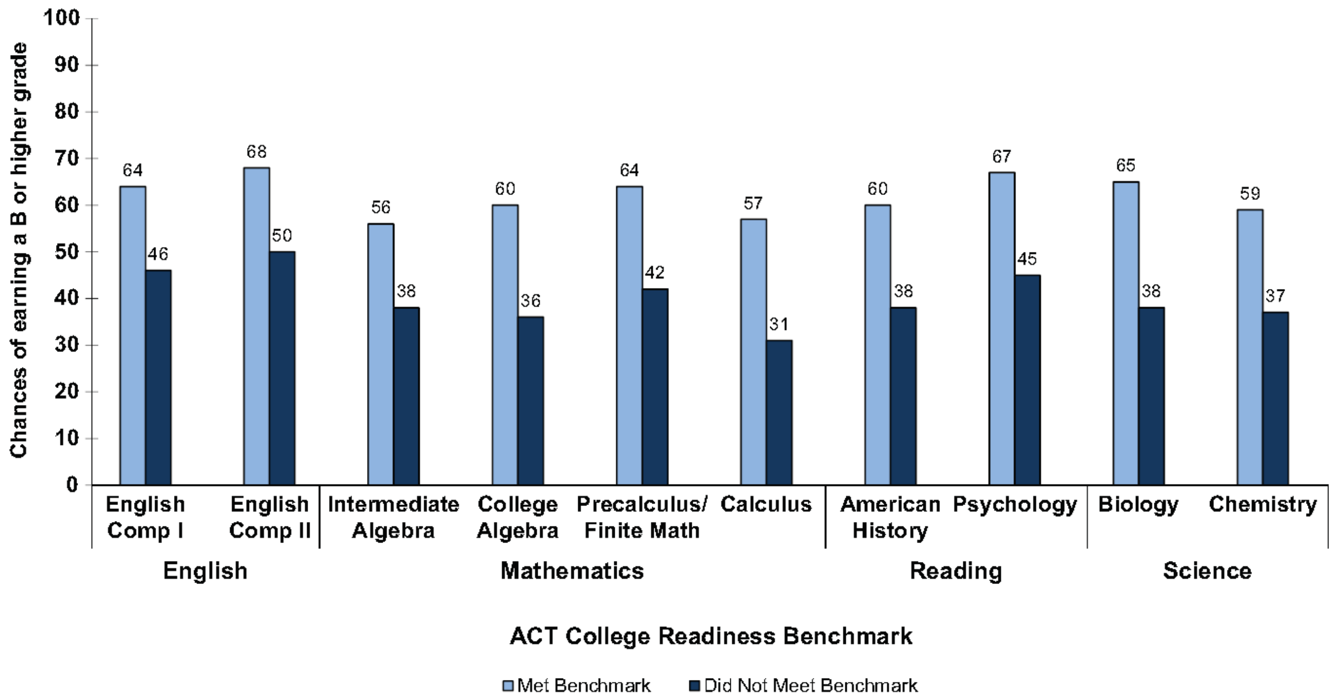

Figure 7.16. Students' chances of earning a B or higher grade in first-year college courses by ACT College Readiness Benchmark attainment at a typical institution

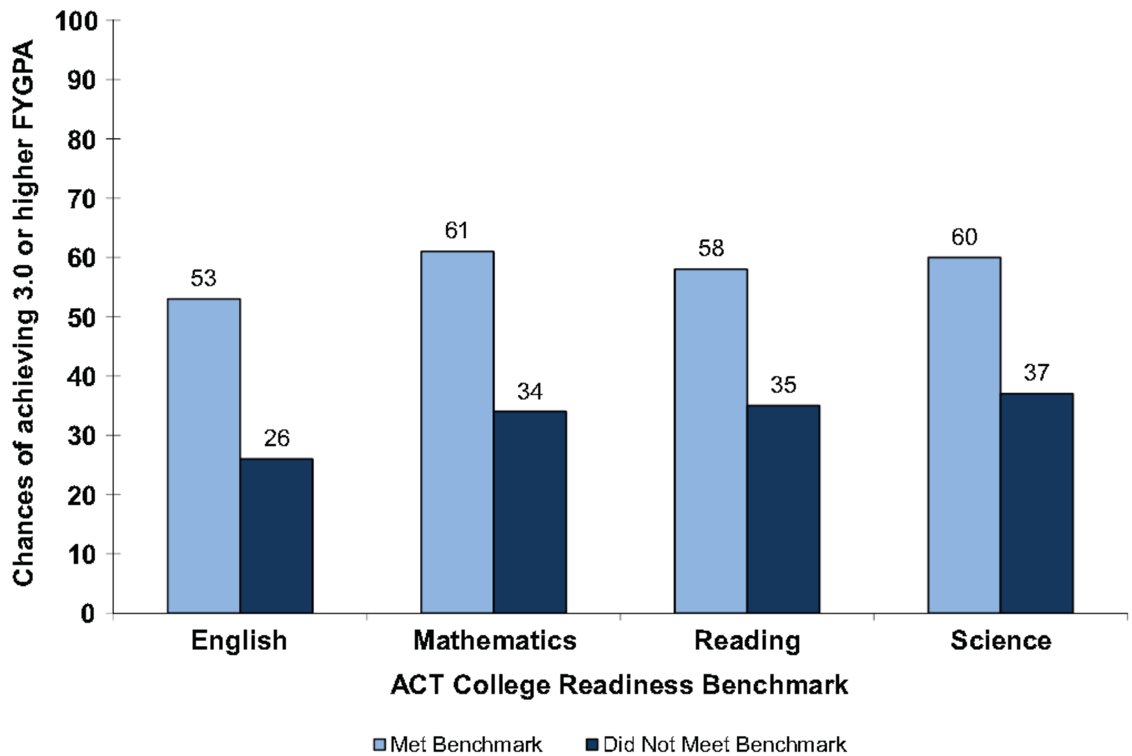

Figure 7.17. Students' chances of achieving a 3.0 or higher FYGPA by ACT College Readiness Benchmark attainment at a typical institution

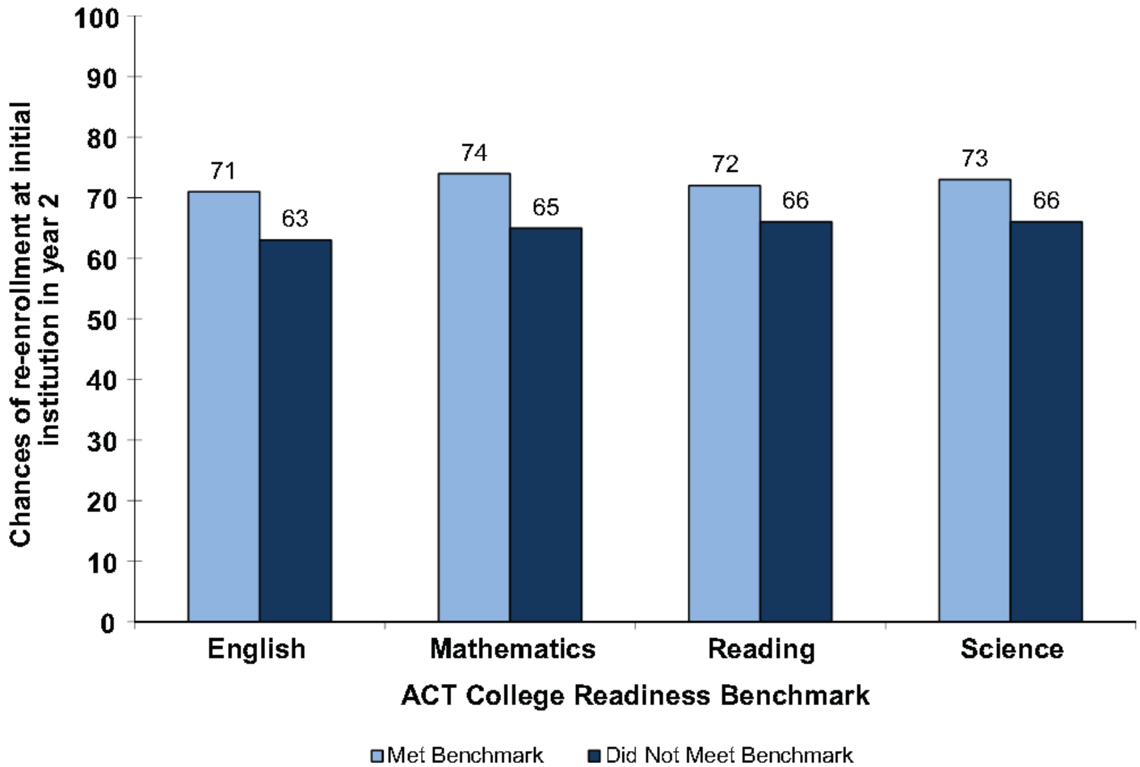

Figure 7.18. Students' chances of remaining enrolled at the initial institution in year two by ACT College Readiness Benchmark attainment

Table 7.25. First-Year College Outcomes by Number of ACT College Readiness Benchmarks Met

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

7.5.2 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 degree completion within three years from the initial institution for students beginning at two-year institutions. Because many students beginning at a two-year institution transfer to a four-year institution without earning an associate degree (Radunzel, 2012), associate'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 7.3.1). The rates were then summarized across institutions.

Results. As shown in Figure 7.19, as ACT Composite score increased, students' chances of completing a degree increased for both two- and four-year students. Additionally, as ACT Composite score increased, two-year students' chances of completing an associate degree or transferring to a four-year institution increased. As an example of the increase for those beginning at a four-year institution, students' chances of completing a bachelor's degree in six years was 41% for those with an ACT Composite score of 20, and it was 67% for those with an ACT Composite score of 30. Higher values of HSGPA were also associated with increased chances of degree completion (see Appendix A of Radunzel and Noble (2012b) for related figures).

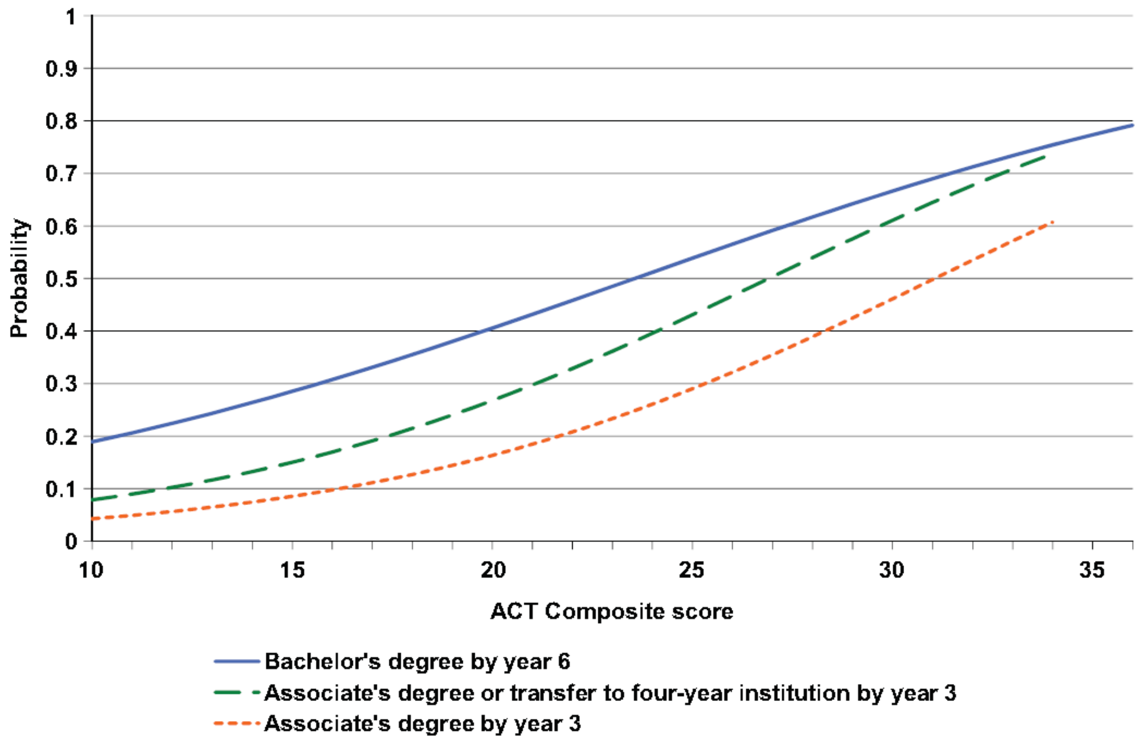

Figure 7.19. Probability of degree completion based on ACT Composite score (Radunzel & Noble, 2012b)

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 7.20 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.

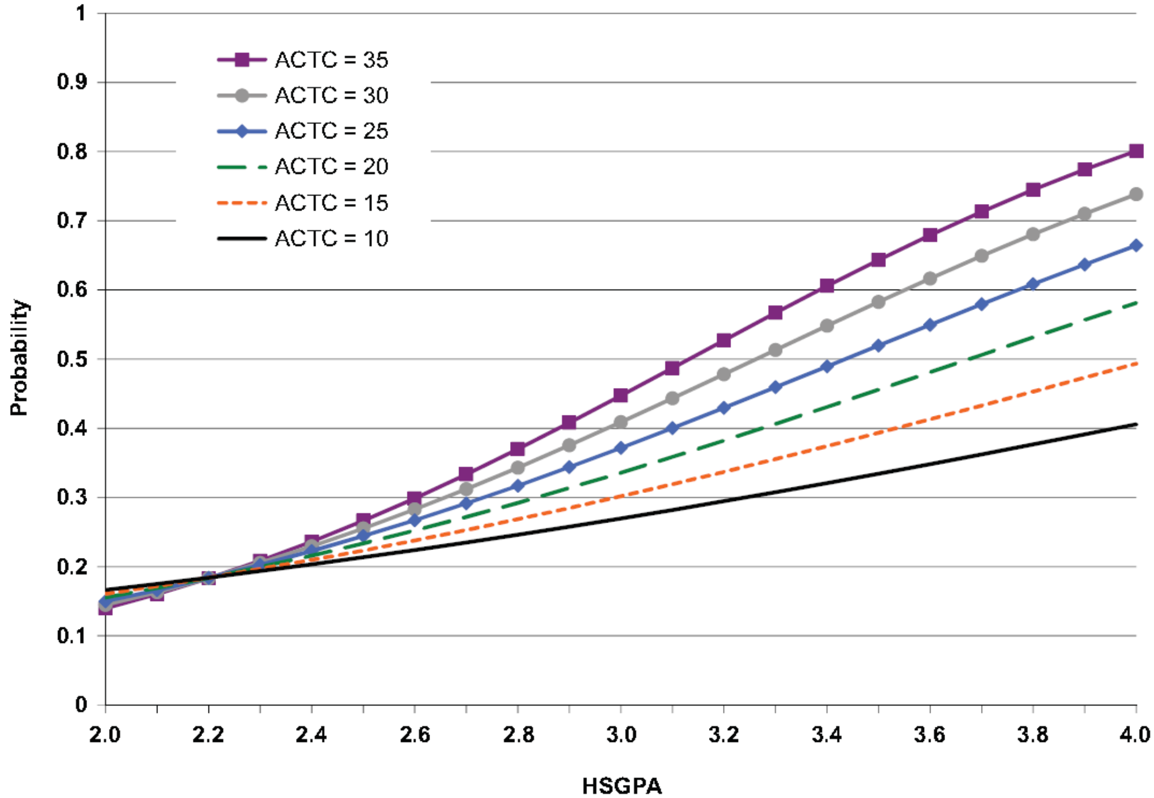

Figure 7.20. Probability of bachelor's degree completion within 6 years, by HSGPA and ACT Composite score (Radunzel & Noble, 2012b)

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).

References

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

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

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

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

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

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

ACT. (2017). *PreACT technical bulletin.* Iowa City, IA: ACT.

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.

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

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

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.

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

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.

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

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.

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.

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: ACT.

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.

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

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.

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.

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.

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.

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.

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

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.

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.

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.

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: ACT.

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.

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.

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.

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.

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.

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.

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.

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.

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

Ree, M. J., Earles, J. A., & Teachout, M. S. (1994). Predicting job performance: Not much more than g. *Journal of Applied Psychology, 79*(4)*,* 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.

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.

Sawyer, R. L. (1985*). Using demographic information in predicting college freshman grades* (ACT Research Report No. 87). Iowa City, IA: ACT.

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.

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.

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: ACT.

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

Schiel, J., & Harmston, M. (2000). *Validating two-stage course placement systems when data are truncated* (ACT Research Report No. 2000-3). Iowa City, IA: ACT.

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: ACT.

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: ACT.

Schiel, J. L., & Valiga, M. J. (2014a). *How ACT scores change with test preparation* (ACT Information Brief 2014-6). Iowa City, IA: ACT.

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.

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.

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. 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: ACT.

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.

Westrick, P., & Allen, J. (2014). *Validity evidence for ACT Compass placement tests (*ACT Research Report No. 2014-2). Iowa City, IA: ACT.

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

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.

Woodruff, D. J., & Ziomek, R. L. (2004b). *Differential grading standards among high schools* (ACT Research Report 2004-2). Iowa City, IA: ACT.

Zhang, Q., & Sanchez, E. I. (2013). *High school grade inflation from 2004 to 2011* (ACT Research Report No. 2013-3). Iowa City, IA: ACT.

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.

Ziomek, R. L., & Svec, J. C. (1995). *High school grades and achievement: Evidence of grade inflation* (ACT Research Report 953). Iowa City, IA: ACT.

Chapter 8

2020 Achievement Summary

Chapter 8: 2020 Achievement Summary

8.1 Student Participation

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

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

Self-reported data describing the ACT examinee population for the 2020 junior class are presented in Table 8.1. A list and count of students' approved accommodations are provided in Table 8.2 These data are based on the 65,443 students who graduated in the spring of 2020 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 8.1 Demographic Characteristics of the 2020 ACT-Tested High School Junior Class

Demographic	%[a]
Gender	
Female	44
Male	46
Other Gender	<1
No response	9
Prefer not to respond	<1
Grade Level When Tested	
Junior	100
Racial-Ethnic Background	
African American/Black	5
White	63
American Indian/Alaska Native	1
Hispanic/Latino	11
Asian	3
Native Hawaiian/Other Pacific Islander	<1
Two or more races	4
Prefer no response/blank	12

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

Table 8.2 List of Approved Accommodations for the 2020 ACT-Tested High School Junior Class

Accommodation Description	N
Approved word-to-word bilingual dictionary/glossary	578
Assistive device: furniture, AAC, switches, adaptive keyboard or mouse	2
Assistive technology: Speech to Text software	42
Auditory amplification/FM system	23
Braille (EBAE, Contracted)	3
Braille (UEB with Nemeth, contracted)	3
Brailled response	2
Breaks as Needed (Standard Time)	287
Computer for writing section response (paper-based testing)	277
Double time on writing section only	52
Double time over multiple days	550
Examinee reads aloud to self in a 1-1 setting	2
Food/drink/medication	188
Home/hospital testing	2
Human reader in 1-1 setting that reads the entire test	256
Large Print (18pt font)	53
Mark answers in test booklet (No Scantron)	39
One and one-half time	5410
One and one-half time over multiple days	566
One-to-one testing	320
Other	175

Accommodation Description	N
Permission to stand during testing	9
Permission to use noise cancelling headphones, white noise machine or listen to instrumental music	6
Preferential Seating	688
Pre-recorded audio (USB)	1496
Printed copy of verbal instructions	18
Raised line drawings (EBAE)	1
Raised line drawings (UEB with Nemeth)	3
Screen reader software for computer-based testing	1
Sign language interpreter for oral instructions only	19
Signing Exact English for entire test	1
Small group testing	6399
Standard time	72
Standard Time over multiple days	98
Text-to-speech	569
Time remaining indicator: countdown timer, note card with time remaining, tap on shoulder	1
Translated test directions (Arabic)	15
Translated test directions (Chinese Mandarin Simp)	7
Translated test directions (Chinese Mandarin Trad)	2
Translated test directions (French)	1
Translated test directions (German)	0
Translated test directions (Hmong Daw)	11
Translated test directions (Japanese)	4
Translated test directions (Portuguese)	1
Translated test directions (Russian)	8
Translated test directions (Somali)	8
Translated test directions (Spanish)	377
Translated test directions (Tagalog)	1
Translated test directions (Vietnamese)	10
Triple time over multiple days	2165
Visual Environment	8
Wheelchair Accessibility	18
Writer/scribe to record responses	1
Writer/scribe to record verbal responses	78

8.2 Student Performance

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 the English Language Arts (ELA) score, a combination of the student's English, reading and writing scores, are reported to the students taking the Wisconsin state test.

8.2.1 Summary statistics, Effective Weights, and Correlations

Summary Statistics

The summary statistics of the ACT test scores for the students taking the primary form of the Wisconsin state test in the 2019–2020 academic year are presented in Table 8.3.

Table 8.3 Summary Statistics of the ACT Test Score Distributions

Primary Form							
Statistic	English	Mathematics	Reading	Science	Composite	Writing	ELA
N	52,614	52,614	52,614	52,614	52,614	52,165	52,165
Mean	19.43	20.41	20.74	20.94	20.5	6.48	19.1
SD	6.28	5.39	6.32	5.27	5.31	1.65	5.34
Skewness	0.41	0.61	0.49	0.18	0.44	–0.13	0.19
Kurtosis	–0.27	–0.55	–0.38	–0.14	–0.41	–0.06	–0.38
Accommodated Form							
Statistic	English	Mathematics	Reading	Science	Composite	Writing	ELA
N	7,684	7,684	7,684	7,684	7,684	7,328	7,328
Mean	13.81	15.69	15.56	15.96	15.37	4.78	13.35
SD	5.02	3.76	5.88	4.71	4.36	1.76	4.96
Skewness	1.59	2.18	1.37	1.2	1.73	0.42	1.07
Kurtosis	2.88	6.15	1.68	2.07	3.21	–0.18	1.09
English Language Learners							
Statistic	English	Mathematics	Reading	Science	Composite	Writing	ELA
N	3,399	3,399	3,399	3,399	3,399	3,352	3,352
Mean	16.46	18.07	17.77	18.59	17.84	5.95	16.49
SD	5.25	4.42	5.32	4.67	4.39	1.60	4.63
Skewness	0.68	1.07	0.74	0.30	0.73	–0.05	0.34
Kurtosis	0.11	0.63	0.42	–0.01	0.14	–0.06	–0.12
Male							
Statistic	English	Mathematics	Reading	Science	Composite	Writing	ELA
N	23,222	23,222	23,222	23,222	23,222	22,915	22,915
Mean	18.9	20.75	20.35	21.01	20.38	6.13	18.38
Std Dev	6.24	5.6	6.42	5.49	5.44	1.67	5.4
Skewness	0.47	0.56	0.51	0.19	0.46	–0.07	0.24
Kurtosis	–0.24	–0.66	–0.39	–0.25	–0.45	–0.18	–0.43
Female							
Statistic	English	Mathematics	Reading	Science	Composite	Writing	ELA
N	23,995	23,995	23,995	23,995	23,914	23,914	23,995
Mean	20.19	21.20	20.98	20.73	6.87	19.90	20.19
Std Dev	5.13	6.16	5.01	5.13	1.53	5.11	5.13
Skewness	0.61	0.48	0.17	0.43	–0.10	0.19	0.61
Kurtosis	–0.53	–0.35	–0.02	–0.35	0.08	–0.30	–0.53
American Indian/Alaska Native							
Statistic	English	Mathematics	Reading	Science	Composite	Writing	ELA

N	403	403	403	403	403	397	397
Mean	15.30	16.88	17.45	17.74	16.97	5.50	15.44
Std Dev	4.76	3.51	4.72	4.21	3.73	1.63	4.38
Skewness	0.55	1.36	0.63	0.04	0.67	–0.10	0.27
Kurtosis	0.00	2.53	0.32	–0.46	0.16	–0.29	–0.17
Asian							
Statistic	English	Mathematics	Reading	Science	Composite	Writing	ELA
N	1,686	1,686	1,686	1,686	1,686	1,677	1,677
Mean	19.61	21.37	21.05	21.28	20.94	6.96	19.83
Std Dev	6.65	5.76	6.51	5.46	5.64	1.66	5.52
Skewness	0.58	0.65	0.59	0.45	0.65	–0.08	0.38
Kurtosis	–0.20	–0.51	–0.29	0.03	–0.19	–0.04	–0.31
Black/African American							
Statistic	English	Mathematics	Reading	Science	Composite	Writing	ELA
N	2,712	2,712	2,712	2,712	2,712	2,651	2,651
Mean	14.07	15.87	16.08	16.08	15.66	5.48	14.59
Std Dev	4.75	3.17	4.80	4.16	3.70	1.64	4.40
Skewness	1.00	1.63	1.01	0.39	1.13	0.05	0.56
Kurtosis	1.16	4.96	1.57	0.62	1.90	–0.37	0.29
Hispanic/Latino							
Statistic	English	Mathematics	Reading	Science	Composite	Writing	ELA
N	5,183	5,183	5,183	5,183	5,183	5,132	5,132
Mean	16.92	18.01	18.68	18.76	18.22	6.22	17.26
Std Dev	5.51	4.26	5.49	4.62	4.45	1.63	4.78
Skewness	0.69	1.20	0.78	0.30	0.83	–0.11	0.39
Kurtosis	0.33	1.11	0.49	0.18	0.51	–0.06	0.05
Native Hawaiian/Other Pacific Islander							
Statistic	English	Mathematics	Reading	Science	Composite	Writing	ELA
N	44	44	44	44	44	43	43
Mean	17.80	19.05	19.59	19.73	19.16	6.51	18.30
Std Dev	5.92	5.44	5.80	5.41	5.19	1.52	4.84
Skewness	0.58	1.05	0.88	0.56	0.84	–0.69	0.46
Kurtosis	–0.23	0.37	0.76	0.27	0.46	0.28	–0.33
White							
Statistic	English	Mathematics	Reading	Science	Composite	Writing	ELA
N	34,380	34,380	34,380	34,380	34,380	34,165	34,165
Mean	20.42	21.27	21.57	21.83	21.40	6.63	19.88
Std Dev	6.09	5.34	6.27	5.06	5.16	1.60	5.16
Skewness	0.34	0.44	0.40	0.15	0.35	–0.14	0.13
Kurtosis	–0.26	–0.76	–0.49	–0.10	–0.47	0.01	–0.37
Two or more races							
Statistic	English	Mathematics	Reading	Science	Composite	Writing	ELA

N	1,866	1,866	1,866	1,866	1,866	1,843	1,843
Mean	19.14	19.77	20.76	20.47	20.15	6.48	19.02
Std Dev	6.13	5.21	6.27	5.19	5.20	1.67	5.30
Skewness	0.45	0.82	0.54	0.20	0.54	–0.06	0.27
Kurtosis	–0.24	–0.20	–0.30	–0.09	–0.29	–0.22	–0.42

Male—Accommodated Form

Statistic	English	Mathematics	Reading	Science	Composite	Writing	ELA
N	4,262	4,262	4,262	4,262	4,262	4,032	4,032
Mean	13.57	15.81	15.33	15.94	15.28	4.49	12.86
Std Dev	4.88	3.90	5.87	4.83	4.39	1.70	4.84
Skewness	1.63	2.23	1.42	1.26	1.79	0.53	1.14
Kurtosis	3.08	6.10	1.84	2.15	3.47	-0.04	1.24

Female—Accommodated Form

Statistic	English	Mathematics	Reading	Science	Composite	Writing	ELA
N	2,549	2,549	2,549	2,549	2,549	2,467	2,467
Mean	14.45	15.68	16.12	16.23	15.74	5.31	14.36
Std Dev	5.30	3.67	6.02	4.63	4.43	1.76	5.12
Skewness	1.47	2.00	1.25	1.07	1.57	0.26	0.98
Kurtosis	2.32	5.27	1.24	1.74	2.39	–0.22	0.82

American Indian/Alaska Native—Accommodated Form

Statistic	English	Mathematics	Reading	Science	Composite	Writing	ELA
N	122	122	122	122	122	118	118
Mean	12.23	14.41	14.46	14.73	14.08	4.26	11.84
Std Dev	3.85	2.52	4.09	3.55	2.87	1.64	4.01
Skewness	1.17	–0.10	1.14	0.77	1.21	0.37	0.90
Kurtosis	2.52	10.06	1.09	1.01	1.44	–0.40	0.18

Asian—Accommodated Form

Statistic	English	Mathematics	Reading	Science	Composite	Writing	ELA
N	342	342	342	342	342	321	321
Mean	12.99	16.43	14.44	16.02	15.07	5.25	13.31
Std Dev	4.24	4.86	4.93	4.41	3.98	1.73	4.26
Skewness	2.27	2.05	1.65	1.20	2.09	0.11	1.10
Kurtosis	7.31	4.17	3.94	2.53	5.74	–0.46	2.02

Black/African American—Accommodated Form

Statistic	English	Mathematics	Reading	Science	Composite	Writing	ELA
N	516	516	516	516	516	475	475
Mean	11.83	14.07	13.27	14.05	13.43	4.34	11.44
Std Dev	3.51	2.20	4.19	3.21	2.74	1.63	3.89
Skewness	2.22	1.51	2.09	1.15	2.56	0.56	1.56
Kurtosis	8.45	9.07	7.03	3.98	10.09	0.00	3.72

Hispanic/Latino—Accommodated Form

Statistic	English	Mathematics	Reading	Science	Composite	Writing	ELA

N	1,245	1,245	1,245	1,245	1,245	1,162	1,162
Mean	12.73	14.79	14.26	14.79	14.27	4.70	12.49
Std Dev	3.94	2.64	4.78	3.77	3.27	1.65	4.11
Skewness	1.93	2.94	1.64	1.39	2.32	0.27	1.08
Kurtosis	5.96	14.85	3.70	3.93	8.02	–0.42	1.75
Native Hawaiian/Other Pacific Islander—Accommodated Form							
Statistic	English	Mathematics	Reading	Science	Composite	Writing	ELA
N	11	11	11	11	11	11	11
Mean	14.27	15.27	16.09	17.73	16.00	5.45	14.55
Std Dev	4.27	1.85	5.84	2.72	3.26	2.46	5.43
Skewness	0.04	0.33	0.11	0.28	0.25	0.24	–0.23
Kurtosis	–0.94	–1.32	–0.90	0.38	–0.04	–0.32	–0.92
White—Accommodated Form							
Statistic	English	Mathematics	Reading	Science	Composite	Writing	ELA
N	3,941	3,941	3,941	3,941	3,941	3,820	3,820
Mean	14.74	16.34	16.65	16.84	16.26	4.92	14.14
Std Dev	5.48	4.11	6.41	5.09	4.80	1.81	5.34
Skewness	1.31	1.88	1.11	1.01	1.40	0.43	0.92
Kurtosis	1.58	3.96	0.65	1.25	1.69	–0.18	0.48
Two or more races—Accommodated Form							
Statistic	English	Mathematics	Reading	Science	Composite	Writing	ELA
N	260	260	260	260	260	247	247
Mean	15.01	16.23	16.67	16.67	16.26	4.82	14.16
Std Dev	5.94	4.32	6.61	5.64	5.16	1.69	5.56
Skewness	1.39	2.04	1.11	1.10	1.44	0.59	0.94
Kurtosis	1.45	4.67	0.41	1.22	1.42	0.06	0.32

Effective Weights

The Composite and ELA scores are the rounded average of the subject test scores. Specifically, the English, mathematics, reading, and science test scale scores are weighted equally to form the Composite 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 and ELA scores, the weights used in the calculation are 1/4 and 1/3 respectively, and they are often referred to as normal 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 test using the formula

$$(effective\ weight)_x = \frac{\Sigma_y cov_{xy}}{\Sigma_x \Sigma_y cov_{xy}},$$

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 the primary test form administered in Wisconsin during the 2019–2020 academic year were computed (see Table 8.4). The effective weight for the English test was computed by adding up the four numbers in the first row. 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. The effective weights for the mathematics, reading, and science were obtained in a similar fashion.

Table 8.5 shows the ranges of effective weights for the Composite and ELA scores based on the primary test forms administered in Wisconsin during the 2019–2020 academic year. 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 and reading tests also contributed to higher effective weights for English and reading in the ELA score.

Table 8.4 Scale Score Covariances for Multiple-Choice Tests from the Primary Test Form

Test	English	Mathematics	Reading	Science
English	39.44	25.92	32.80	26.34
Mathematics	25.92	29.04	23.60	22.71
Reading	32.80	23.60	40.00	25.67
Science	26.34	22.71	25.67	27.76

Table 8.5 Effective Weights of the ACT Tests from the Primary Test Form

Test	Composite	ELA
English	0.28	0.36
Mathematics	0.22	--
Reading	0.27	0.36
Science	0.23	--
Writing	--	0.29

Correlations

Table 8.6 shows the correlations among the ACT test scores based on operational data from the primary test form administered in Wisconsin during the 2019–2020 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 6.

Table 8.6 Correlations Among the ACT Test Scores

Score	English	Mathematics	Reading	Science	Composite	Writing	ELA
English	1.00	0.76	0.82	0.79	0.93	0.55	0.92
Mathematics		1.00	0.69	0.80	0.88	0.47	0.74
Reading			1.00	0.77	0.91	0.51	0.90
Science				1.00	0.91	0.50	0.80
Composite					1.00	0.56	0.93
Writing						1.00	0.78
ELA							1.00

8.3 Detailed Performance Description

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

8.3.1 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 highlighting students' relative strengths and areas for improvement on each subject. The reporting category scores replaced the subscores (e.g., Intermediate Algebra/Coordinate Geometry) that were reported previously.

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 score for the reporting category corresponding to the Benchmark on the overall English test. Students with percent correct scores at or above the minimum percent correct score 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 2, 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).

8.3.2 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 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 2

Table 8.7 presents the summary statistics of writing domain scores and the overall writing scores based on the primary writing test form administered in Wisconsin during the 2019–2020 academic year. Table 8.8 presents the correlations among these scores.

Table 8.7 Summary Statistics of the ACT Writing and Writing Domain Scores

Statistic	Ideas & Analysis	Development & Support	Organization	Language Use & Conventions	Writing Score
N	52,165	52,165	52,165	52,165	52,165
Mean	6.42	5.93	6.33	6.80	6.48
SD	1.75	1.73	1.72	1.52	1.65
Skewness	–0.21	0.02	–0.25	–0.10	–0.13
Kurtosis	–0.05	–0.31	–0.10	0.25	–0.06

Table 8.8 Correlations among the ACT Writing and Writing Domain Scores

Score	Ideas & Analysis	Development & Support	Organization	Language Use & Conventions	Writing Score
Ideas & Analysis	1.00	0.93	0.98	0.94	0.98
Development & Support		1.00	0.94	0.89	0.94
Organization			1.00	0.92	0.98
Language Use & Conventions				1.00	0.96
Writing Score					1.00

References

Powers, Li, Suh, and Harris (2016). *ACT Reporting Category Interpretation Guide Version 1.0*. (ACT Working Paper No. 2016-05). Iowa City, IA: ACT.

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

Chapter 9

College and Career Readiness Standards and College Readiness Benchmarks

Chapter 9: College and Career Readiness Standards and College Readiness Benchmarks

9.1 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 tables of College and Career Readiness Standards can be found in Chapter 8 of the ACT® Technical Manual beginning on page 8.18. Additional information can be found at the ACT website (http://www.act.org/content/act/en/college-and-career-readiness/standards.html).

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.

9.2 ACT's College and Career Readiness Standards

9.2.1 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.

9.2.2 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.

9.2.3 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 10 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 10 forms of the ACT mathematics test. There are 60 items in each ACT mathematics test form, so 600 ACT mathematics items were reviewed in all. An illustrative table displaying the information provided to the mathematics content team for one ACT mathematics test form is shown in Table 9.1.

The shaded areas in Table 9.1 show the items that met the .80-or-above item difficulty criterion for each of the score ranges. As illustrated in Table 9.1, 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 9.2 reports the total number of test items reviewed for each content area.

These procedures 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.
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.

Table 9.1 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
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 9.2 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

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 pre-selected 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 9.3. 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 9.3 Percentage of Agreement of 1997 National Expert Review

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

9.2.4 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.

9.2.5 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 the Standards reflect:

- the most important knowledge and skills for college and career readiness,
- what is being measured by the items in each score range, and
- 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.

9.2.6 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.

9.3 ACT's College Readiness Benchmarks

9.3.1 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 9.4). 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 9.4 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 (two-year and four-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.

9.3.2 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 9.4). 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 (two-year, four-year less selective, and four-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 9.5 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 9.6; 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).

Table 9.5 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 9.6 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 is based on National Student Clearinghouse data.

Summary

Students, parents, and counselors can use the Benchmarks to determine the academic areas in which students are ready for college coursework 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).

9.3.3 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 coursework 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 9.1). 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).

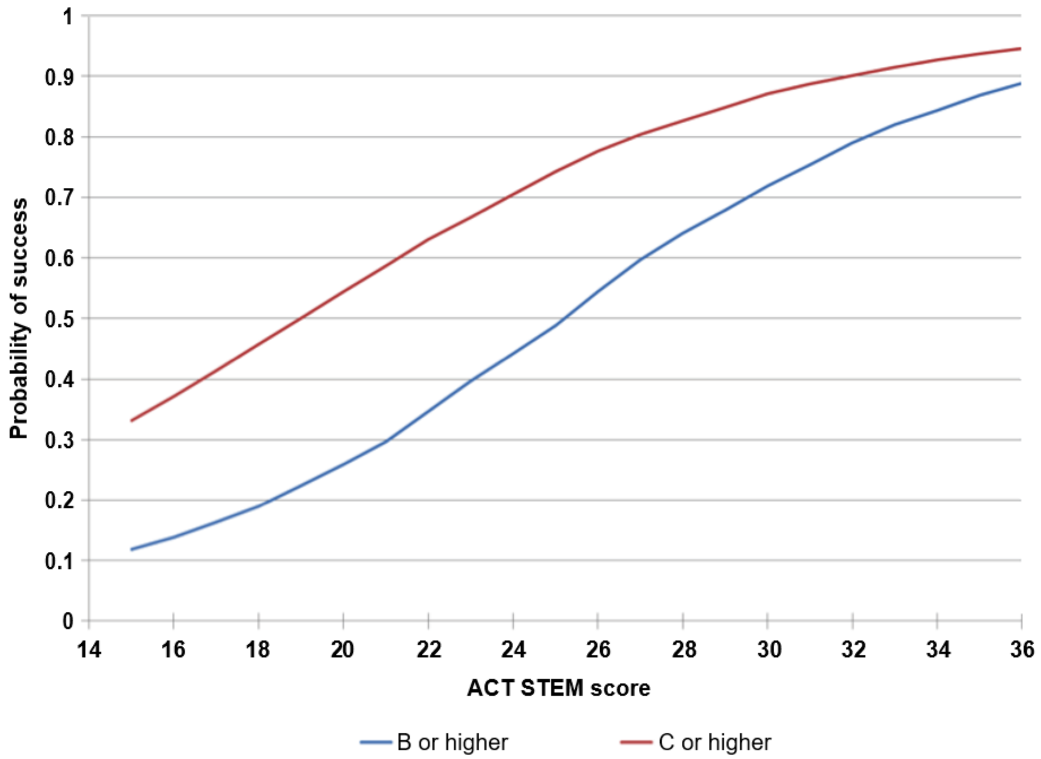

Figure 9.1 The typical probability of success in STEM-related courses by the ACT STEM score. The mathematics course was Calculus I. The science courses included Biology, Chemistry, Physics, and Engineering.

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.

9.3.4 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. A concordance table was used to convert students' ACT writing scores to current ACT writing scores that were then 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 for 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 9.2). This cutoff score was also associated with an approximate 75% chance of earning a C or higher grade.

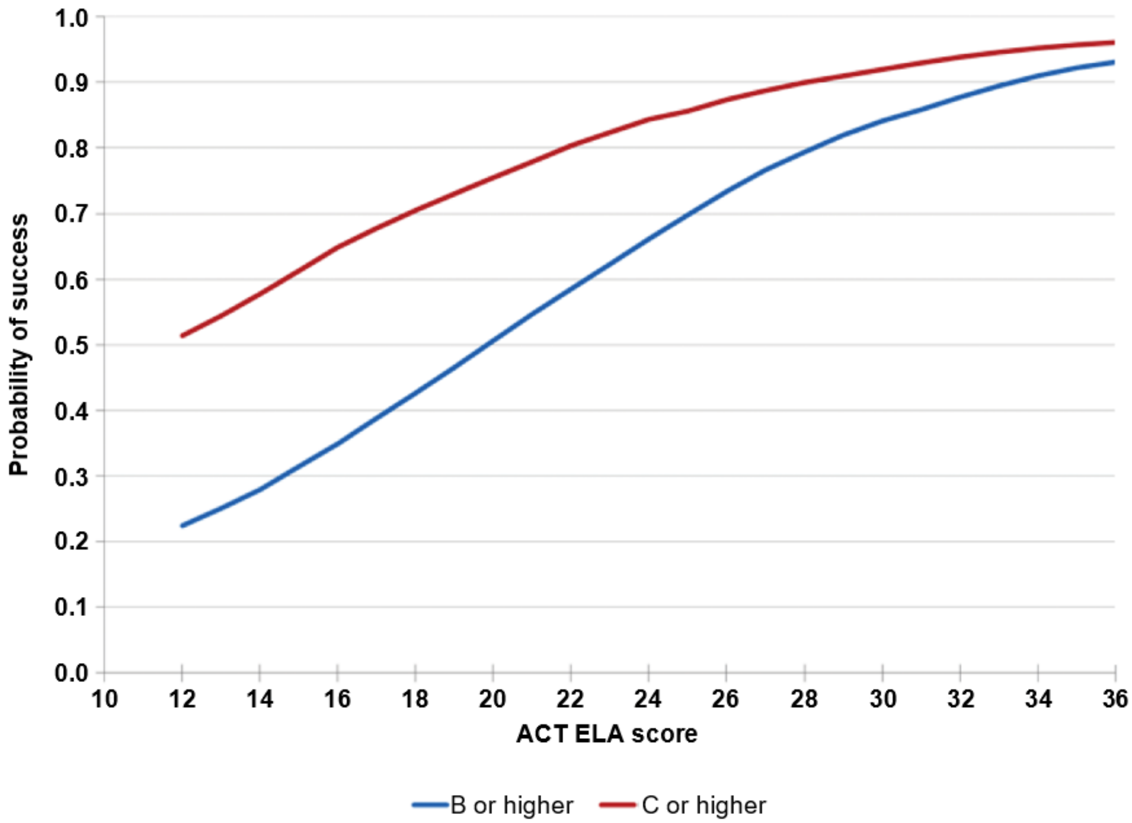

Figure 9.2 The typical probability of success in ELA-related courses by the ACT ELA score. The English related course was English Composition I. The social science courses included American History, Other History, Psychology, Sociology, Political Science, and Economics.

Summary

The ACT ELA Readiness Benchmark can be used to help gauge overall student readiness for ELA-related coursework. In the academic year 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 coursework 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.

9.3.5 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.

9.3.6 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.

References

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

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

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

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

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

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

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.

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.

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.

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.

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.

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.

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: ACT.

Chapter 10

Other ACT Components

Chapter 10: Other ACT Components

10.1 The ACT Interest Inventory

10.1.1 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.

10.1.2 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. 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 10.1) 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.

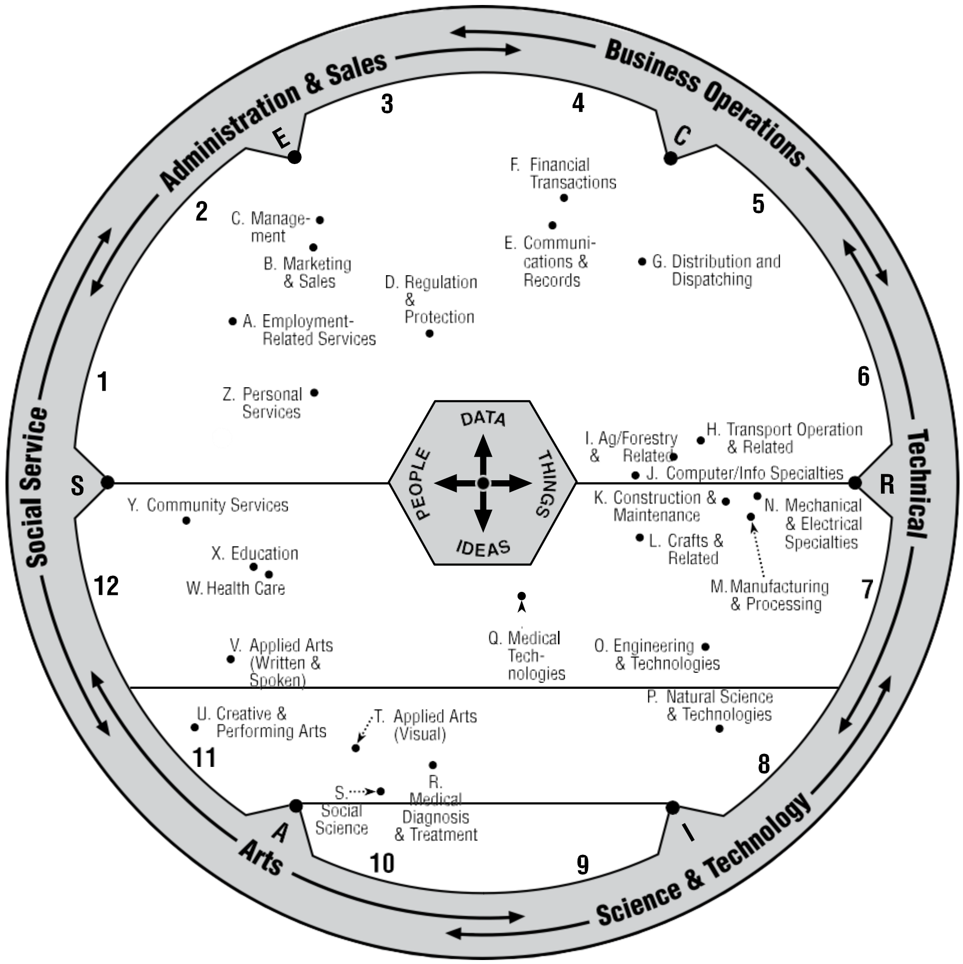

Figure 10.1 The ACT Career Map

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 10.1, 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).

10.1.3 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).

10.1.4 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. Data from a large national sample were used to establish three levels of fit 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 of *The ACT Technical Manual*).

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, the 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).

10.2 Profile Reports

Annual profile reports are based on national and state-specific data to provide various overviews of national and state data. The profile reports are organized into an executive summary covering five-year trends, academic achievement with score distributions and results by demographic groupings, college readiness and curriculum results, scores by student-reported career and educational aspirations, and results of the optional writing test. An example of the table of contents is provided.

Table of Contents

Section I: Executive Summary — **Page 5**
- Average Composite Scores: 5 Years of Testing
- Percent Meeting 3 or 4 Benchmarks: 5 Years of Testing
- Percent Meeting STEM Benchmark: 5 Years of Testing
- Percent Taking A Core Curriculum: 5 Years of Testing
- Five Year Trends—Percent of Students Who Met College Readiness Benchmarks
- Five Year Trends—Average ACT Scores
- Five Year Trends—Average ACT Scores by Level of Preparation
- Five Year Trends—Percent and Average Composite Score by Race/Ethnicity
- Five Year Trends—Achievement in STEM
- Proficiency Toward Understanding Complex Text

Section II: Academic Achievement — **Page 11**
- Average ACT Composite Scores by Race and Core Curriculum Status
- ACT Score Distributions, Cumulative Percents, and Averages
- Subject Area Reporting Categories
- Average ACT Composite Scores for Race/Ethnicity by Level of Preparation
- Average ACT Scores by Race/Ethnicity
- Average ACT Composite Scores for Gender by Level of Preparation
- Average ACT Scores by Gender
- ACT Score Quartile Values

Section III: College Readiness & Impact of Course Rigor — **Page 17**
- Percent of Students Meeting 3 or 4 College Readiness Benchmarks by Core College Curriculum Status
- Percent of Students in College and Career Readiness Standards (CCRS) Score Ranges
- Percent of Students Who Met ACT College Readiness Benchmark Scores by Gender
- Percent of Students Who Met ACT College Readiness Benchmark Scores by Race/Ethnicity
- Likely ACT National Career Readiness Certificate (NCRC) Level Based Upon ACT Composite Score
- College Readiness Benchmark Percent and Average ACT Scores by Overall High School Curriculum
- College Readiness Benchmark Percent and Average ACT Scores by Content-Specific Curriculum
- College Readiness Benchmark Percent and Average ACT Scores by Common Course Patterns
- College Readiness Benchmark Percent and Average ACT Scores for Gender by Common Course Patterns

Section IV: Career and Educational Aspirations — **Page 25**
- Average ACT Composite Scores by Race and Student Postsecondary Aspirations
- Distribution of Planned Educational Majors for All Students by College Plans
- Average ACT Composite Scores for Racial/Ethnic Groups by Post-Secondary Educational Aspirations
- Students' Score Report Preferences at Time of Testing

Section V: Optional Writing Test Results — **Page 31**
- Average ACT Writing Scores by Race/Ethnicity
- Average ACT English Language Arts Constituent Scores by Race/Ethnicity and Gender for Students Who Took ACT Writing
- Average ACT English Language Arts Outcomes by Race/Ethnicity and Gender for Students Who Took ACT Writing

Figure 10.2 Profile report table of contents (retrieved from *https://www.act.org/content/dam/act/unsecured/documents/2020/2020-National-ACT-Profile-Report.pdf*)

10.3 Growth Modeling

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 these: 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 section, 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.

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, the ACT most directly supports gain-based and conditional status models.

10.3.1 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 a 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.

10.3.2 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:

$$\text{Predicted ACT Mathematics Score} = \beta 0 + \beta 1 \times \text{ACT Plan English Score} + \beta 2 \times \text{ACT Plan Mathematics Score} + \beta 3 \times \text{ACT Plan Reading Score} + \beta 4 \times \text{ACT Plan Science Score} + \beta 5 \times \text{Months Elapsed}$$

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 include 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.

10.4 ACT Retesting

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) investigated 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 the 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 10.3, 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 10.4 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. When grade levels were used as a proxy for curriculum coverage and with 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.

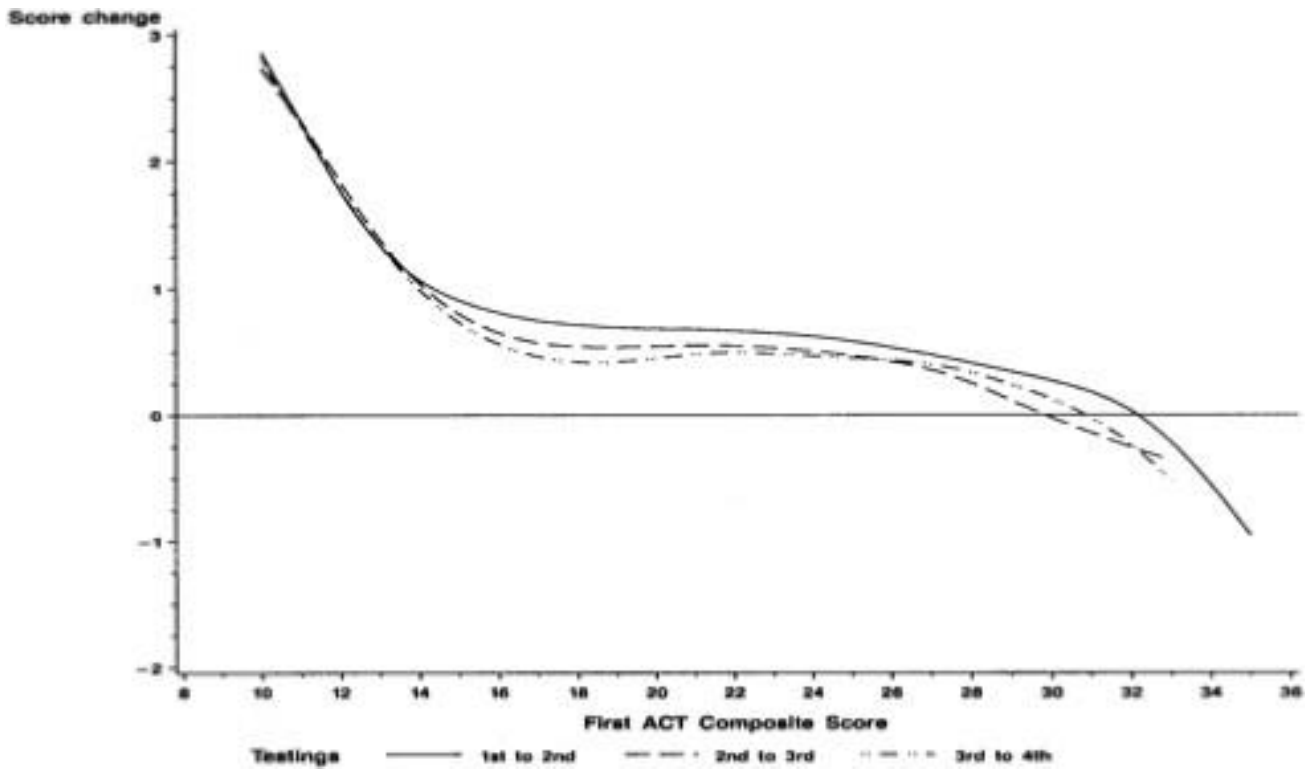

Figure 10.3 Changes in Composite test scores from 1st to 2nd, 2nd to 3rd, and 3rd to 4th testing

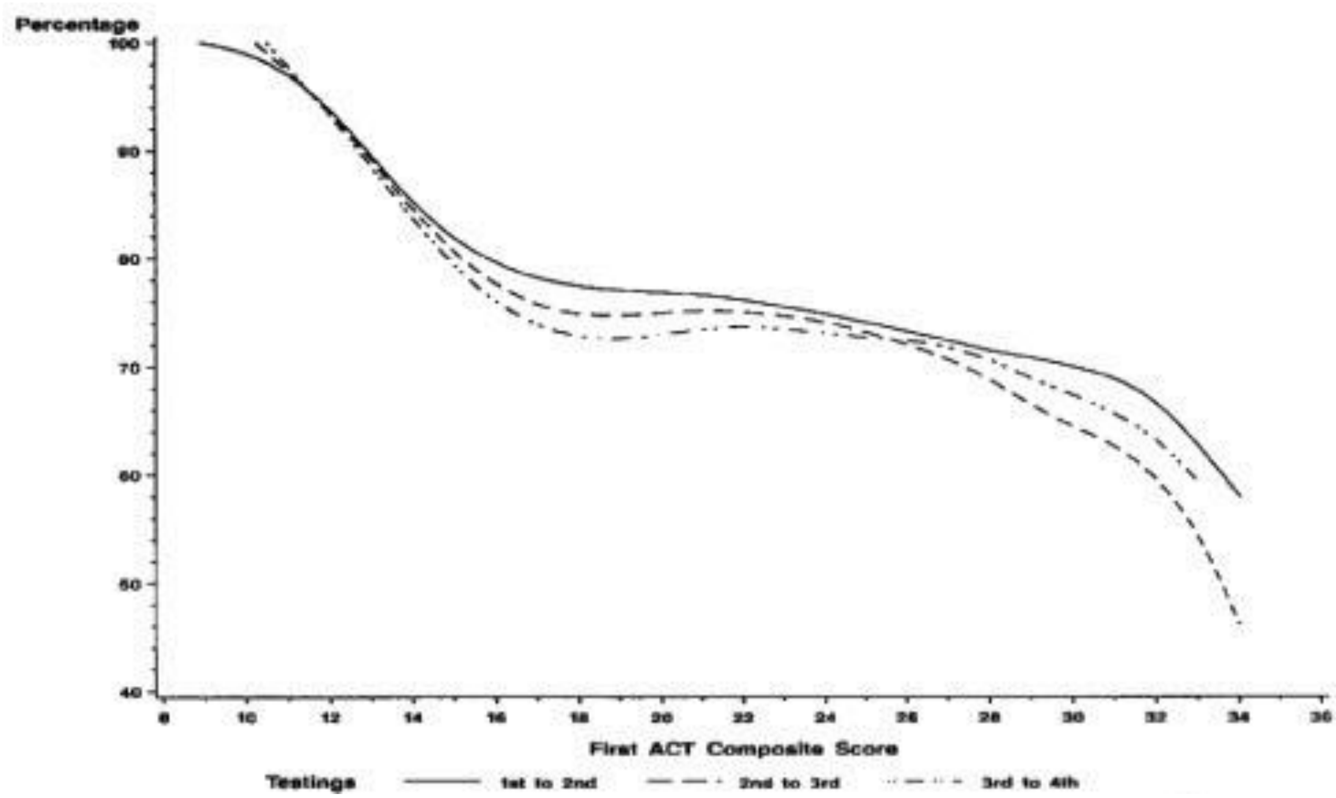

Figure 10.4 Percentage of students maintaining or increasing score from 1st to 2nd, 2nd to 3rd, and 3rd to 4th testing

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

Table 10.1 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. (2009). *The ACT interest inventory technical manual.* Iowa City, IA: ACT. 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: ACT.

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

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.

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.

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.

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

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

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